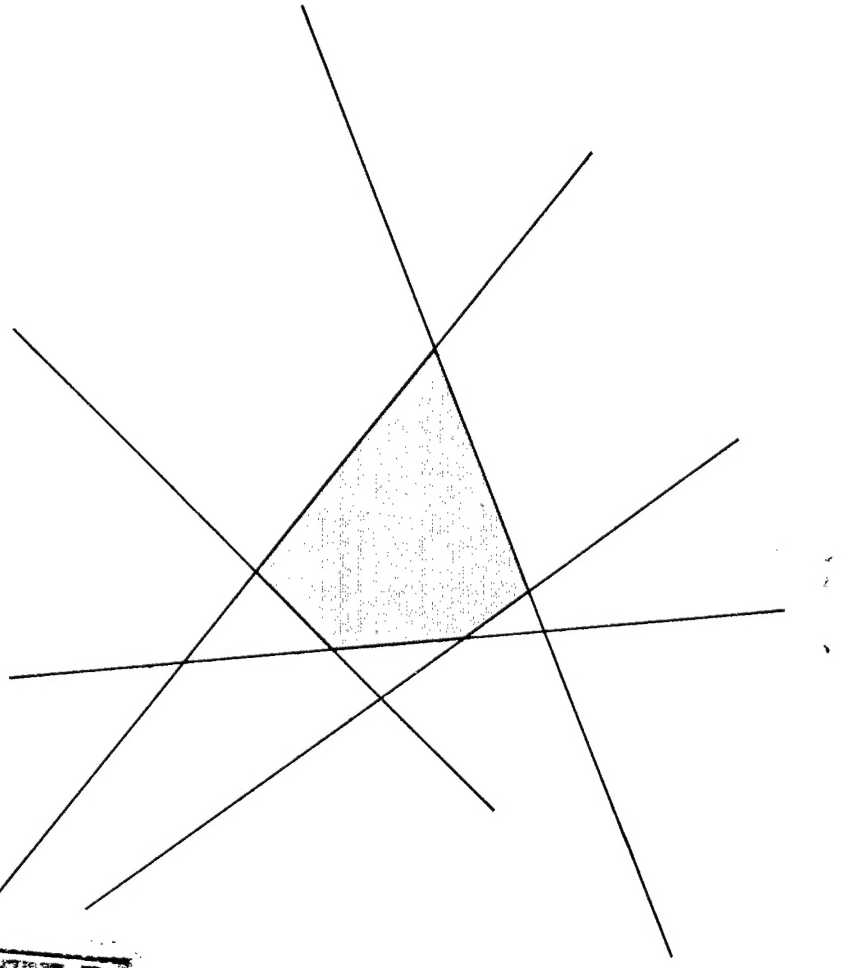
ORC 76-16
JUNE 1976

THR[RIA-77-U1084
IN C. .EORY

by
WILLIAM S. JEWELL



UNIVERSITY OF CALIFORNIA · BERKELEY

THREE PAPERS IN CREDIBILITY THEORY[†]

THE USE OF COLLATERAL DATA IN CREDIBILITY THEORY: A HIERARCHICAL MODEL

by

William S. Jewell

BAYESIAN REGRESSION AND CREDIBILITY THEORY

by

William S. Jewell

BAYESIAN INVERSE REGRESSION AND DISCRIMINATION: AN APPLICATION OF CREDIBILITY THEORY

by

Rudolph Avenhaus and William S. Jewell

JUNE 1976

ORC 76-16

[†]Research was supported by the International Institute for Applied Systems Analysis, Laxenburg, Austria, and the reproduction was supported by the U. S. Army Research Office - Research Triangle Park under Grant DAAG29-77-G-0040.

THE FINDINGS IN THIS REPORT ARE NOT TO BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, UNLESS SO DESIGNATED BY OTHER AUTHORIZED DOCUMENTS.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ORC 76-16	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* THREE PAPERS IN CREDIBILITY THEORY		5. TYPE OF REPORT & PERIOD COVERED Research Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR*(s)* William S. Jewell		8. CONTRACT OR GRANT NUMBER*(s)* DAAG29-77-G-0040
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P-14240-M
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE June 1976 13. NUMBER OF PAGES 98
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT *(of this Report)* Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)* Bayesian Statistics Credibility Theory Regression Calibration Hierarchical Models		
20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)* (SEE ABSTRACT)		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

FOREWORD

During 1974-1975, the author spent his sabbatical leave at the International Institute for Applied Systems Analysis, Laxenburg, Austria, where he was able to continue his research in credibility methods in a stimulating international scientific community.

Because of the difficulty of obtaining copies of research memoranda published during that period, it seems desirable to reproduce them in this format for distribution to interested colleagues, sponsors, and students. Naturally, credit for support and initial distribution of this work should remain with IIASA; two of the papers have been submitted to journals for possible publication.

RM - 75 - 24

INTERNATIONAL INSTITUTE FOR IIASA APPLIED SYSTEMS ANALYSIS

RESEARCH MEMORANDUM

THE USE OF COLLATERAL DATA IN CREDIBILITY THEORY: A HIERARCHICAL MODEL

William S. Jewell

June 1975

SCHLOSS LAXENBURG
2361 Laxenburg
AUSTRIA

RM-75-24

THE USE OF COLLATERAL DATA IN CREDIBILITY THEORY: A HIERARCHICAL MODEL

William S. Jewell

June 1975

The Use of Collateral Data in Credibility Theory: A Hierarchical Model

William S. Jewell*

Abstract

In classical credibility theory, a linearized Bayesian forecast of the fair premium for an individual risk contract is made using prior estimates of the collective fair premium and individual experience data. However, collateral data from other contracts in the same portfolio is not used, in spite of intuitive feelings that this data would contain additional evidence about the quality of the risk collective from which the portfolio was drawn. By using a hierarchical model, one makes the individual risk parameters exchangeable, in the sense of de Finetti, and a modified credibility formula is obtained which uses the collateral data in an intuitively satisfying manner. The homogeneous formula of Bühlmann and Straub is obtained as a limiting case when the hyperprior distribution becomes "diffuse".

0. Introduction

In the usual collective model of risk theory [1], the random variables generated by individual risks are assumed to be independent, once the individual risk parameters are known. However, a priori, only collective (portfolio) statistics are available, taken from a distribution which is mixed over a prior distribution of the parameter. We assume that unlimited statistics are available for the collective as a whole, and a limited amount of experience (sample) data for individual risks drawn at random from the collective.

---

*University of California, Berkeley, and International Institute for Applied Systems Analysis, Laxenburg, Austria

In classical credibility theory, we make a linearized Bayesian forecast of the next observation of a particular individual risk, using his experience data and the statistics from the collective; the resulting formula, which has been known in various forms for over fifty years, requires only the individual sample mean, and the first and second moments from the collective.

If one attempts to use collateral data from other risks in a credibility forecast of a certain individual risk, it turns out that this cohort data has zero weight, and is discarded in favor of the assumed-known collective statistics. This is essentially because the various individual risk parameters are assumed to be independent and representative samples from the prior distribution.

This result is disturbing to many analysts, who feel that data from other risks in the portfolio contains valuable collateral information about the collective. In several of their models, Bühlmann and Straub [3,4] argue that, since the (mixed) moments of the collective must be estimated anyway, a credibility forecast should be only in terms of cohort data. They achieve a partial result of this kind by using a proportional function of all experience data; this forces the use of cohort data into an estimate of the collective mean, but the second moment components are still required. In [12], the author describes a model in which the individual risk parameters were correlated through an "externalities" model; the resulting formula uses both cohort sample data and the first

and second collective moments. In [18], Taylor describes a model in which the "manual premium" (collective mean) is itself a random variable, and also obtains a formula in which collateral data is used. Finally, we should mention that similar arguments are advanced about the use of cohort data in the otherwise unrelated "empirical Bayes" models [14, 16].

In this paper, we attempt a reconciliation of these approaches, based upon the ideas of hierarchical models [13,14, 15] and model identification [17,19]. Although we obtain results similar to those already described in [12], the justification is completely different, and, we believe, provides a more natural explication of the situations in which collateral data should be used.

1. The Basic Model

In the basic model of the collective, we imagine that individual risk contracts are characterized by a risk parameter, θ, which is drawn from a known prior density, $p(\theta)$. A cohort, or portfolio, of such contracts consists of a finite population $[\theta_1,\theta_2,\ldots,\theta_r]$, whose members are drawn independently from the same density.

Then, given θ_i, we suppose that we have likelihood densities, $p_i(x_{it}|\theta_i)$,[1] which govern the generation of n_i independent

---

[1]We adopt the usual convention that all densities are indicated by $p(.)$, the arguments indicating the appropriate random variable(s). The random variables, themselves, are indicated where necessary by a tilde. Finally, to avoid complicated

(continued)

and identical realizations of the risk random variable, $\tilde{x}_{it}(t = 1,2,\ldots,n_i)$. In other words, from the total portfolio, we have r individual experience data records, $\underline{x}_i = [x_{i1},x_{i2},\ldots,x_{in_i}]$, which, together, we refer to as the total experience, X. Note that each process is stationary over time, but that we (temporarily) permit the individual risks to have different distributions. In particular, we need to define the first two conditional moments:

$$m_i(\theta_i) = \mathscr{E}\{\tilde{x}_{it}|\theta_i\} \quad ; \qquad v_i(\theta_i) = \mathscr{V}\{\tilde{x}_{it}|\theta_i\} \ . \tag{1.1}$$

Prior to the data, $p(\theta)$ is the same prior density for any arbitrary risk drawn from the collective; thus, a priori, we have the following average moments for risks of the i^{th} and j^{th} types:

$$m_i = \mathscr{E}\{\tilde{x}_{it}\} = \mathscr{E}\{m_i(\tilde{\theta}_i)\} \quad ; \tag{1.2}$$

$$E_{ij} = \mathscr{E}\mathscr{E}\{\tilde{x}_{it};\tilde{x}_{ju}|\tilde{\theta}_i;\tilde{\theta}_j\} = \begin{cases} 0 & (i \neq j) \\ \mathscr{E}\{v_i(\tilde{\theta}_i)\} & (i = j) \end{cases} \tag{1.3}$$

$$D_{ij} = \mathscr{C}\{m_i(\tilde{\theta}_i);\ m_j(\tilde{\theta}_j)\} = \begin{cases} 0 & (i \neq j) \\ \mathscr{V}\{m_i(\tilde{\theta}_i)\} & (i = j) \end{cases} \tag{1.4}$$

---

1 (cont'd) subscripts, we define the multiple conditional expectation:

$$\mathscr{E}\mathscr{E}\mathscr{E}\{f(\tilde{a},\tilde{b},\tilde{c},)|\tilde{b}|\tilde{c}\}$$

as being the expectation of $f(a,b,c)$ using measure $p(a|b,c)$, followed by the expectation using measure $p(b|c)$, followed by the expectation using $p(c)$. Any of these arguments may be multiple, and other operators, such as variance, $\mathscr{V}$, and covariance, $\mathscr{C}$, may be used.

Note in particular that there are no covariances between risks i and j ≠ i for two reasons:

(i) assumed independence between $\tilde{x}_{it}$ and $\tilde{x}_{ju}$, given θ_i and θ_j;

(ii) assumed independence between $\tilde{\theta}_i$ and $\tilde{\theta}_j$.

The total prior-to-data covariance between individual risks is then:

$$\mathscr{C}\{\tilde{x}_{it};\tilde{x}_{ju}\} = \begin{cases} E_{ii} + D_{ii} & (i = j, t = u) \\ D_{ii} & (i = j, t \neq u) \\ 0 & (i \neq j) \end{cases} . \qquad (1.5)$$

The basic problem of credibility theory is to forecast the next observation, $\tilde{x}_{s,n_s+1}$, of a <u>selected risk</u>, s, given the total data from all risks, $X = [x_i | (i = 1,2,\ldots,r)]$, and using the linear function:

$$f_s(X) = a_o + \sum_{i=1}^{r} \sum_{t=1}^{n_i} a_{it}\, x_{it} , \qquad (1.6)$$

in which the coefficients $(a_o; a_{it})$ are chosen so as to approximate the conditional mean $\mathscr{E}\{\tilde{x}_{s,n_s+1} | X\}$ in the least-squares sense, over all prior possible data records, p(X).

The appropriate least-squares formulae have been presented elsewhere (see, e.g., [7,12]). It turns out, for the basic model described above, that:

(i) $a_{it} = a_i$ $(i = 1,2,\ldots,r)(t = 1,2,\ldots,n_i)$ because of the stationarity assumption;

(ii) $a_i = 0$ $(i \neq 0,s)$ because $D_{sj} = 0$ $(j \neq s)$, that is, $\tilde{\theta}_j$ and $\tilde{\theta}_s$ are independent.

Defining the i^{th} credibility factor, Z_i, and time constant, N_i, as:

$$Z_i = n_i/(n_i + N_i) \quad ; \quad N_i = E_{ii}/D_{ii} \quad ; \tag{1.7}$$

and the i^{th} experience sample mean, $\bar{x}_i$, as:

$$\bar{x}_i = \frac{1}{n_i} \sum_{t=1}^{n_i} x_{it} \quad , \tag{1.8}$$

we obtain the final credibility forecast as:

$$f_s(X) = (1 - Z_s)m_s + Z_s\bar{x}_s + O(x_{i\neq s,t}) \quad . \tag{1.9}$$

Various interesting interpretations of this classical result are possible [7,8,12], and it is known that (1.9) is, in fact, the exact Bayesian conditional mean for a large and important class of prior and likelihood densities [9,10].

2. Objections and Previous Results

Two practical objections to the result (1.9) seem to be raised in the literature. The first is that three prior-to-data moments, m_s, E_{ss}, and D_{ss}, must be estimated from the collective for each risk which is forecast. Even in the more usual, identical-risk case, where $m_i = m$, $E_{ii} = E$, and $D_{ii} = D$, for all samples $i = 1,2,\ldots,r$, (1.9) provides no assistance in estimating the common moments. This concern is related to the second objection, namely, that there ought to be some use for the cohort data, $\{x_{i\neq s,t}\}$, since it is precisely from this data that one would attempt to form estimates of the first and second moments in actual practice. This collateral data ought,

then, to be used either to form initial estimates of m, E, and D, or, in the case in which one had vague prior estimates of them, to somehow revise them as more portfolio-wide data becomes available. Notice that we are not talking about any problems of non-stationarity, such as inflation, or shifts in the risk environment, but just the vague notion that our collective might, in some way, be different from the initially-assumed statistics.

Bühlmann and Straub [3] were the first to point out that one can force all the data in X to be used by setting a_o in (1.6) equal to zero, and constraining the remaining coefficients to give a forecast which is unbiased, as in (1.9). For the simple model of the last section, in which the $\tilde{x}_{it}$ are not identically distributed, we obtain:

$$f_s(X) = (1 - Z_s) \left\{ m_s \frac{\sum_{i=1}^{r} \left(\frac{m_i}{D_{ii}}\right) Z_i \bar{x}_i}{\sum_{j=1}^{r} \left(\frac{m_j^2}{D_{jj}}\right) Z_j} \right\} + Z_s \bar{x}_s \quad . \qquad (2.1)$$

The term in braces, which used all the sample data, even that of risk s, is a substitute for m_s in (1.9); however, there is no simplification as far as collective moments to be estimated are concerned, since all the m_i, E_{ii}, and D_{ii} are used.

But in the important case where all risks are assumed to be identically distributed, for the same value of θ, (2.1) simplifies to:

$$f_s(X) = (1 - Z_s) \left\{ \frac{\sum_{i=1}^{r} Z_i \bar{x}_i}{\sum_{j=1}^{r} Z_j} \right\} + Z_s \bar{x}_s \quad , \tag{2.2}$$

and now the forecast depends upon $Z_i = n_i/(n_i + N)$, with $N = E/D$ as a ratio between variance components which must be estimated from the collective. Of course, the forecast (2.2) must give a higher value to the mean-square error which was used to find (1.9).

If all data records are of the same length, $n_i = n$ and $Z_i = Z = n/(n + N)$, $(i = 1,2,\ldots,r)$, the surrogate for m_s in the braces in (2.2) becomes simply:

$$\sum_{i=1}^{r} \bar{x}_i/r = \sum_{i=1}^{r} \sum_{t=1}^{n} x_{it}/rn \quad , \tag{2.3}$$

the grand sample mean of all cohort data!

In some work on "related risk" models [12], the author assumed a situation in which the risk parameters $\underline{\tilde{\theta}} = [\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_r]$ are statistically <u>dependent</u>, with known joint prior. The only effect of this assumption is to introduce non-zero terms into the last line of (1.4), viz.:

$$D_{ij} = \mathscr{C}\{m_i(\tilde{\theta}_i);\ m_j(\tilde{\theta}_j)\} \tag{2.4}$$

for all i,j. If the underlying risk likelihoods are different, then a multidimensional credibility model [7,11] must be used with an $r \times r$ system of equations solved to find a matrix of credibility factors. However, in the important special case

where the risks are identically distributed, given $\underline{\theta}$, $p(\underline{\theta})$ consists of exchangeable random variables, and there are only four collective moments, m, E, and, say, D_{11} and D_{12} for the cases in which $i = j$ and $i \neq j$, respectively, in (2.4). One may easily show that, with this correlation between risk parameters added, (1.9) becomes:

$$f_s(X) = (1 - Z_s) \left\{ \frac{(D_{11} - D_{12})m + D_{12} \sum_{i=1}^{r} Z_i \bar{x}_i}{(D_{11} - D_{12}) + \sum_{j=1}^{r} Z_j} \right\} + Z_s \bar{x}_s \quad , \tag{2.5}$$

where the credibility factors now require a modified correlation time constant, N_{12}:

$$Z_i = n_i/(n_i + N_{12}) \quad ; \quad N_{12} = E/(D_{11} - D_{12}) \quad . \tag{2.6}$$

As in (2.2), the expression in braces in (2.5) is an estimate for the mean m_s, which can be seen to be different from m, because of the non-representative way in which the cohort of r risks may have been selected. As the correlation between the parameters vanishes, $D_{12} \to 0$, $D_{11} \to D$, and (2.5) reduces to the usual formula (1.9), with all the collateral data being thrown away.

Although this model is satisfactory from the mathematical point of view of explaining when cohort data would be used in a linear forecast, it does not show why there could be correlation in the collective, why the risk parameters should be exchangeable random variables, and under what conditions this correlation would be weak or strong. For this purpose, we need to extend the traditional model of the collective into a hierarchical model.

3. A Hierarchical Model

In our expanded model, the concepts of individual risk random variables, risk parameters, and a cohort of risks chosen from a collective are retained, but we imagine that our collective, the one under study, is not necessarily representative of other possible collectives which are drawn from some larger universe of collectives.

Formally, this means that there is a collective selection hyperparameter, $\tilde{\varphi}$, which describes how possible collectives may vary from one another, when chosen from some hyperprior density $p(\varphi)$. Once φ is chosen and the collective characteristics are defined, then the risk parameters $[\theta_i]$ are chosen for each of the r members of our cohort, independently, and identically distributed from a prior density $p(\theta|\varphi)$. Finally, the n_i experience samples for each individual risk i are drawn independently from a likelihood, $p_i(x_{it}|\theta_i,\varphi)$. Notice that the risk parameters and the individual risks are now independent only if φ is given; from the prior-to-selection-of-collective point of view, there is apparent correlation between cohort results because of the mixing on φ.

This somewhat abstract model has a very practical interpretation. Imagine an insurance company in which the individual risk is an individual insurance contract, and the collective is just a portfolio of similar coverages within our company. It is well recognized that portfolios vary from company to company, depending upon sales strategy, available customers, local risk conditions, etc.; our portfolio may be better or

worse, than, say, the nationwide average. The universe of collectives, then, corresponds to the union of all possible risk contracts of this type in the nation, for which we may assume adequate statistics are available. Thus, in a hierarchical model, we hope to use nationwide statistics, together with all the data from our portfolio, not only to predict next year's fair premium for individual risks, but also to draw inferences about what kind of a portfolio we have.

For the development of a least-squares forecast, we start with the individual risk moments of $p(x_{it}|\theta_i,\varphi)$:

$$m_i(\theta_i,\varphi) = \mathcal{E}\{\tilde{x}_{it}|\theta_i,\varphi\} \quad ; \quad v_i(\theta_i,\varphi) = \mathcal{V}\{\tilde{x}_{it}|\theta_i,\varphi\} \quad , \tag{3.1}$$

and, from the usual conditional arguments, form the universal-average mean of the i^{th} type:

$$M_i = \mathcal{E}\{\tilde{x}_{it}\} = \mathcal{E}\mathcal{E}\{m_i(\tilde{\theta}_i,\tilde{\varphi})|\tilde{\theta}_i|\tilde{\varphi}\} \quad . \tag{3.2}$$

The universal covariances, using the conditional independence properties described above, are:

$$\mathcal{C}\{\tilde{x}_{it};\tilde{x}_{ju}\} = \begin{cases} F_{ii} + G_{ii} + H_{ii} & (i = j, t = u) \\ G_{ii} + H_{ii} & (i = j, t \neq u) \\ H_{ij} & (i \neq j) \quad , \end{cases} \tag{3.3}$$

where

$$F_{ii} = \mathcal{E}\mathcal{E}\{v_i(\tilde{\theta}_i,\tilde{\varphi})|\tilde{\theta}_i|\tilde{\varphi}\} \quad , \tag{3.4}$$

$$G_{ii} = \mathcal{E}\mathcal{V}\{m_i(\tilde{\theta}_i,\tilde{\varphi})|\tilde{\theta}_i|\tilde{\varphi}\} \quad , \tag{3.5}$$

and

$$H_{ij} = \mathscr{C}\left\{\mathscr{E}\{m_i(\tilde{\theta}_i,\tilde{\varphi})\,|\,\tilde{\varphi}\} \;;\; \mathscr{E}\{m_j(\tilde{\theta}_j,\tilde{\varphi})\,|\,\tilde{\varphi}\}\right\} \quad . \tag{3.6}$$

Several remarks are in order. From one point of view, what we have done is to introduce correlation between risk parameters of members of the same collective, for on comparing the above with (1.5) as modified by (2.4), we get the formal equivalences:

$$E_{ii} \equiv F_{ii} \;;\; D_{ii} \equiv G_{ii} + H_{ii} \;;\; D_{ij} \equiv H_{ij} \quad (i \neq j) \quad . \tag{3.7}$$

However, the interpretation is completely different, as we have seen.

The second observation is that is might seem worth while to decouple the $\tilde{x}_{it}$ from $\tilde{\varphi}$, and make the likelihood only dependent upon $\tilde{\theta}_i$; this might simplify some of the computations above, but does not diminish the number of individual prior-to-selection-of-collective moments needed.

However, in the important special case where the individual risk contracts are similar, giving identical likelihoods, given θ_i and φ, it can be seen that only four moments remain: M, F, G, and H. These may be interpreted in terms of our simpler model by noticing that it is as if the moments of Section 1 had a hidden dependence upon an unknown parameter φ. Calling those moments, then, $m(\varphi)$, $E(\varphi)$, and $D(\varphi)$, we see that the universal moments are equivalent to:

$$M = \mathscr{E}m(\tilde{\varphi}) \;;\; \mathscr{E}F = \mathscr{E}E(\varphi) \;;\; G = \mathscr{E}D(\varphi) \;;\; H = \mathscr{V}m(\tilde{\varphi}) \quad . \tag{3.8}$$

In other words, M, F, and G are universe-averaged versions of our previous m, E, and D. H, however, is new, and represents the variance of the fair premium over all possible collectives.

4. Universal Forecasts

Continuing with the important special case of identical risk distributions, it follows easily from least-squares theory and the above definitions that the optimal credibility forecast for the hierarchical model is:

$$f_s(X) = (1 - Z_s) \left\{ \frac{GM + H \sum_{i=1}^{r} Z_i \bar{x}_i}{G + H \sum_{i=1}^{r} Z_j} \right\} + Z_s \bar{x}_s \qquad (4.1)$$

where now a new universal time constant, N_U, appears in the credibility factors:

$$N_U = F/G \quad ; \quad Z_i = n_i/(n_i + N_U) \quad . \qquad (4.2)$$

Alternatively, we can get (4.1) from (2.5) and (3.7).

Following an idea of Taylor for his model [18], we note that (4.1) can be split into two parts:

$$f_s(X) = (1 - Z_s) \hat{M}(X) + Z_s \bar{x}_s \quad ; \qquad (4.3)$$

$$\hat{M}(X) = \frac{GM + H \sum_{i=1}^{r} Z_i \bar{x}_i}{G + H \sum_{j=1}^{r} Z_j} \quad . \qquad (4.4)$$

The second formula may be regarded as a revision of the "prior expected manual premium", M, using the experience data of all

members of the cohort to obtain an "adjusted manual premium", $\hat{M}(X)$. This revised manual premium is then used in an ordinary credibility formula with the appropriate individual credibility factor, Z_s, for the forecast risk s.

The credibility revision of the universal mean (4.4) depends in a complicated manner upon the amount of data from each risk. However, if all data records are of the same length n, then $Z_i = Z = n/(n + N_U)$ for all i, and (4.4) can be rewritten:

$$M(X) = (1 - Z_C)M + Z_C\left(\frac{1}{r}\sum_{i=1}^{r}\bar{x}_i\right) \quad , \tag{4.5}$$

where the collective credibility factor, Z_C, is:

$$Z_C = \frac{rnH}{F + nG + rnH} = \left[\frac{rH}{G + rH}\right]\left[\frac{n}{n + (F/G + rH))}\right] \quad . \tag{4.6}$$

If rH is large compared to G, this function increases at first more rapidly than the common individual credibility factor Z, as n increases; however, Z_C has an asymptotic limit less than unity, so that (4.5) is not a credibility formula in the usual sense; that is, the grand sample mean is not ultimately "fully credible" for $m(\varphi)$.

This puzzling result can be explained by remembering that the risk parameters of the cohort $[\theta_i | i = 1,2,\ldots,r]$, once picked, remain the same for all n. Therefore, if one estimates a fair premium for an arbitrary new member of the portfolio, say, with risk parameter θ_{r+1}, then there remains the possibility that the cohort sample is biased. Thus Z_C does not approach

unity with increasing n, unless rH >> G, which means that a large enough portfolio contains a representative sample of risk parameters. This effect is not important in our estimate of $\tilde{x}_{s,n+1}$ because of the factor $(1 - Z_s)$ in (4.1).

If, on the other hand, we did wish to estimate the fair premium averaged over the current portfolio:

$$\mathcal{E}\left\{\frac{1}{r}\sum_{i=1}^{r}\tilde{x}_{i,n+1}|X\right\} \quad ,$$

then one can show that (4.5) is still correct if a different credibility factor,

$$Z_o = (nG + rnH)/(F + nG + rnH) \quad , \tag{4.7}$$

is used; this does approach unity with increasing n.

5. Limiting Cases

The time constant $N_U = F/G$ is just the universe-average version of the classical Bühlmann time constant $N = E/D$, so that (4.3) is in a certain sense similar to (1.9). However, the factor $H = \mathcal{V}m(\tilde{\varphi})$ is completely new, and it is interesting to examine limiting cases.

If $H \to 0$, then we may say that all collectives are representative samples from the rather narrow universe of collectives in which there is little variance in fair premium. Thus, $M \to m$, $G \to D$, $N_U \to N$, and $Z_C \to 0$. No updating of the fair premium is necessary from the collateral data, and (4.3)-(4.4) reduce to the classical model (1.9).

On the other hand, if $H \to \infty$, this means that collectives are drastically different from one another, or in Bayesian language, we have a "diffuse prior" on $m(\tilde{\varphi})$. Then from (4.4) or (4.6), we see that, whenever there is cohort data, it is "fully credible" for $m(\tilde{\varphi})$, and (4.1) reduces to the Bühlmann-Straub proportional forecast (2.2)!

The same effect occurs in (4.6) as $r \to \infty$, but for a different reason: the grand sample mean of X is almost surely the correct mean, $m(\varphi)$, for our collective, and thus M is eliminated.

6. Approximation Error

The value of any forecast must be judged in terms of the mean-square error:

$$I = \mathscr{E}\left\{[\tilde{x}_{s,n_s+1} - f_s(\tilde{X})]^2\right\} \quad . \tag{6.1}$$

A certain portion of this error is due to individual fluctuation, and cannot be removed; the remainder is essentially an approximation error between the chosen forecast and the optimal Bayesian forecast, $\mathscr{E}\{\tilde{x}_{s,n_s+1}|X\}$. (See, e.g., [12].) We now examine the mean-square error for several of the forecasts suggested previously.

The first and simplest possibility is to take the universal mean, $f_s(X) = M$, as an estimator. Then:

$$I_1 = F + G + H \quad , \tag{6.2}$$

that is, no component of variance is removed.

The second possibility, suggested by the surrogate for the collective mean in (2.2), is to take the credibility-weighted mean of all cohort data, $f_s(X) = \Sigma Z_i \bar{x}_i / \Sigma Z_j$, giving:

$$I_2 = F + G\left[1 + \frac{(1 - 2Z_s)}{\Sigma Z_j}\right] , \tag{6.3}$$

which removes the fluctuation component H, but may increase the second term for $Z_s < \frac{1}{2}$.

A third collective-wide possibility which has already been justified is the "adjusted manual premium", $\hat{M}(X)$, in (4.4), for which:

$$I_3 = F + G + H\left[\frac{G(1 - 2Z_s)}{G + H\Sigma Z_j}\right] . \tag{6.4}$$

Turning now to forecasts which use the data from the individual risk in a special way, we could use the Bühlmann-Straub homogenous formula (2.2), giving:

$$I_4 = F + G(1 - Z_s)\left[1 + \frac{(1 - Z_s)}{\Sigma Z_j}\right] . \tag{6.5}$$

Also of interest would be an individual forecast in which the cohort data is ignored, (1.9):

$$I_5 = F + G(1 - Z_s) + H(1 - Z_s)^2 . \tag{6.6}$$

Finally, we have the variance when the optimal universal forecast (4.1) is used:

$$I_6 = F + G(1 - Z_s) + H\left[\frac{G}{G + H\Sigma Z_j}\right](1 - Z_s)^2 . \tag{6.7}$$

Notice that none of the forecasts removes F; this is the irreducible variance component. Comparison of different forecasts depends in general upon the values of G, H, and the credibility factors; for example, one cannot say that I_2 is uniformly better than I_1.

The following relationships do hold, however, for all values of the coefficients:

$$I_6 < I_3 < I_2 \quad ;$$

$$I_6 < I_4 < I_2 \quad ;$$

$$I_6 < I_5 < I_1 \quad .$$

This effectively removes I_1 and I_2 from the second-rank contenders, after the optimal forecast I_6.

The Bühlmann-Straub formula, I_4, would seem to have special appeal because of the fact that H is removed completely. However, $I_6 < I_4$ always; and when $H \to \infty$, I_6 approaches a finite limit as well. Conversely, the classical individual credibility mean-square error, I_5, continues to increase as the universal prior becomes more diffuse, and this is the basic justification for including the cohort data.

7. Normal Hierarchical Family

A special case of interest is when all densities discussed in Section 3 are normal. If N(a,b) refers to the normal density with mean a and variance b, then by setting:

$$p(x_{it}|\theta_i,\varphi) = N(\theta_i,F) \quad ; \quad p(\theta_i|\varphi) = N(\varphi,G) \quad ; \quad p(\varphi) = N(M,H) \quad , \tag{7.1}$$

we find that the universal forecast (4.1) is exactly the Bayesian conditional mean $\mathscr{E}\{\tilde{x}_{s,n_s+1}|X\}$.

Further, the adjusted manual premium, $\hat{M}(X)$ (4.4), is $\mathscr{E}\{\tilde{\varphi}|X\}$. The joint distribution $p(\underline{\theta}|X)$, as well as $p(\phi|X)$, are both normal, and their precision matrices may be found by elementary calculations.

8. Related Work

A linear Bayesian model which is hierarchical in form has been given by Lindley and Smith [13,14,15]. In this model, $\underline{\tilde{x}}$, $\underline{\tilde{\theta}}$, and $\underline{\tilde{\varphi}}$ are random vectors for which $\mathscr{E}\{\underline{\tilde{x}}|\underline{\tilde{\theta}},\underline{\tilde{\varphi}}\} = A_1\underline{\tilde{\theta}}$, and $\mathscr{E}\{\underline{\tilde{\theta}}|\underline{\tilde{\varphi}}\} = A_2\underline{\tilde{\varphi}}$, A_1 and A_2 being matrices of appropriate dimension. The underlying distributions are all assumed to be multinormal, with $\mathscr{E}\{\underline{\tilde{\varphi}}\}$ and the covariances assumed to be known constants. When specialized to our model, results similar to Section 7 are obtained.

In [18], Taylor develops a credibility model in which the "manual premium", m, is revised according to "the average actual claim amount per unit risk in the entire collective in the year of experience". His assumptions are different from ours, in that m "has a prior distribution at the beginning of the year of experience", but "for fixed m, each $m(\theta_i)$ is fixed" (in our notation). I interpret this as saying, in effect, that there is a hidden parameter, φ, which is still left in $m = m(\varphi)$, after averaging over the θ_i. However, I have been unable to further relate the two models, and his formulae have the disadvantage that, as "the prior distribution on m" becomes

degenerate, his forecast does not reduce to the usual credibility formula.

9. Conclusion

In conclusion, we mention that our hierarchical model implies that the joint distribution of the risk parameters at the level of the insurance company is:

$$p(\underline{\theta}) = \int \prod_{i=1}^{r} p(\theta_i | \varphi)\ p(\varphi)\ d\varphi \quad ,$$

which is equivalent to assuming that the risk parameters are exchangeable random variables. This powerful concept, due to de Finetti [5,6], is a natural modelling assumption for problems in which a random sample generates a finite population whose members are distinguishable only by their indices, as in our selection of a portfolio from an abstract collective. [14], Section 6, and [15] contain further discussions of the applicability of exchangeability. In a certain sense, what our model does is to use exchangeability to introduce correlation among the cohort θ_i, in the same way that a Bayesian prior introduces correlation among successive individual samples. In both cases, this prior correlation vanishes as the actual values of φ and $\underline{\theta}$ become identified.

G. Ferrara once asked how credibility experience rating could be used in a company where there are no prior statistics. By referring the prior estimation problem to a higher level of data collection, and by using all the experience data generated

by the company's contracts as one learns about the actual portfolio quality, we believe that the model developed here goes a long way towards answering this question.

References

[1] Bühlmann, H. Mathematical Methods in Risk Theory, Springer-Verlag, Berlin, 1970.

[2] Bühlmann, H. "Experience Rating and Credibility", ASTIN Bulletin, 4, Part 3, (July, 1967), pp. 199-207.

[3] Bühlmann, H. and Straub, E. "Glaubwürdigkeit für Schadenssätze", Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker, 70 (1970), pp. 111-113. Trans. C.E. Brooks, "Credibility for Loss Ratios," ARCH 1972.2.

[4] Bühlmann, H. "Credibility Procedures", Sixth Berkeley Symposium on Mathematical Statistics, (1971), pp. 515-525.

[5] De Finetti, B. Probability, Induction, and Statistics, J. Wiley, New York (1972), p. 266.

[6] De Finetti, B. Theory of Probability, Vol. II, J. Wiley, New York (to appear).

[7] Jewell, W.S. "Multi-Dimensional Credibility", ORC 73-7, Operations Research Center, University of California, Berkeley, (April, 1973). To appear, Journal of Risk and Insurance.

[8] Jewell, W.S. "The Credible Distribution", ORC 73-13, Operations Research Center, University of California, Berkeley, (August, 1973). ASTIN Bulletin, VII, Part 3 (March, 1974), pp. 237-269.

[9] Jewell, W.S. "Credible Means are Exact Bayesian for Simple Exponential Families", ORC 73-21, Operations Research Center, University of California, Berkeley, (October, 1973). ASTIN Bulletin, VIII, Part 3 (September, 1974), pp. 77-90.

[10] Jewell, W.S. "Regularity Conditions for Exact Credibility", ORC 74-22, Operations Research Center, University of California, Berkeley, (July, 1974). To appear, ASTIN Bulletin.

[11] Jewell, W.S. "Exact Multidimensional Credibility", ORC 74-14, Operations Research Center, University of California, Berkeley, (May, 1974). Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker, Band 74, Heft 2 (1974), pp. 193-214.

[12] Jewell, W.S. "Model Variations in Credibility Theory", ORC 74-25, Operations Research Center, University of California, Berkeley, (August, 1974). To appear Proceedings of Actuarial Research Conference on Credibility Theory, Berkeley, Sept., 1974. Academic Press, New York.

[13] Lindley, D.V. "Bayesian Least Squares", Bull. Inst. Internat. Statist., 43, 2, pp. 152-153.

[14] Lindley, D.V. Bayesian Statistics, A Review, Regional Conferences Series in Applied Mathematics, SIAM, Philadelphia, 83 pp.

[15] Lindley, D.V. and Smith, A.F.M. "Bayes Estimates for the Linear Model", Jour. Roy. Statist. Soc., B, 34, (1972), pp. 1-41.

[16] Maritz, J.S. Empirical Bayes Methods, Methuen, London (1970), 159 pp.

[17] Smallwood, R.C. "A Decision Analysis of Model Selection", IEEE Trans. on Systems Science and Cybernetics, SSC-4, 3 (Sept. 1969), pp. 333-342.

[18] Taylor, G.C. "Experience Rating with Credibility Adjustment of the Manual Premium", ASTIN Bulletin, 7, Part 3 (March, 1974).

[19] Wood, E.F. "A Bayesian Approach to Analyzing Uncertainty Among Stochastic Models", RR-74-16, International Institute for Applied Systems Analysis, Laxenburg, Austria (Sept., 1974).

RM - 75 - 63

INTERNATIONAL INSTITUTE FOR IIASA APPLIED SYSTEMS ANALYSIS

RESEARCH MEMORANDUM

BAYESIAN REGRESSION AND CREDIBILITY THEORY

William S. Jewell

November 1975

SCHLOSS LAXENBURG
A-2361 AUSTRIA

RM-75-63

BAYESIAN REGRESSION AND CREDIBILITY THEORY

William S. Jewell

November 1975

Abstract

The development of a Bayesian theory of regression requires special distributional assumptions and rather complicated calculations. In this paper, general formulae for predicting the mean values of the regression coefficients and the mean outcomes of future experiments are developed using the methods of credibility theory, a linearized Bayesian analysis originally used in actuarial problems. No special distributional assumptions on prior or error distributions are needed, and heteroscedastic errors in both the dependent and independent variables are permitted. The first group of formulae hold for arbitrary design matrices and dimensionality of input, since, as common in Bayesian methods, there are none of the usual problems of identifiability. However, in the event that the design matrix has full rank, the credibility results are equivalent to a linear mixture of the prior mean prediction and the classical (generalized) least-squares regression predictor; thus, the credibility result provides a bridge between full Bayesian methods and classical estimators. One can also find easily the preposterior covariance matrix for the credibility estimators, and it is shown that prior information and the results from prior experiments can be cascaded in a particularly intuitive manner. Many special applications of the credibility formulae are possible because of the generality of the assumptions.

Bayesian Regression and Credibility Theory

William S. Jewell

Introduction

Regression theory plays a fundamental role in statistical model-building, parameter estimation, and forecasting. In recent years, the need to incorporate prior information into these models has stimulated the development of Bayesian methods of regression analysis, particularly in the field of econometrics [8,20,21,22,24,32]. However, the resulting formulae are usually complex, and require quite stringent assumptions on the error likelihoods and on the prior distributions of parameters.

Credibility theory, which was developed for a variety of simple predictive problems in insurance [4,5,12,13,14,15,17], is a linearized Bayesian method for forecasting mean values which circumvents many of the difficulties of a full Bayesian analysis; furthermore, in many cases of practical interest, the simplified formulae are also exact. In this paper, which was stimulated by the initial work of Hachemeister and Taylor [10,25], we apply credibility ideas to the full range of Bayesian regression models.

1. Classical Multiple Regression

In the classical model of linear normal multiple regression [8,23], we assume that an n×1 random vector of observable *output variables*, $\tilde{y}$, satisfies the linear model

$$\tilde{y} = X\beta + \tilde{u} \tag{1.1}$$

where X is a known n×k matrix of observations on k independent variables, called the *data* or *design matrix*, β is a k×1 vector of unknown *regression coefficients*, and $\tilde{u}$ is an n×1 random vector of unobservable *error variables*. If we assume that $\tilde{u}$ is multinormally distributed, with zero mean and known covariance matrix C,

$$\mathscr{C}\{\tilde{u};\tilde{u}\} = \mathscr{C}\{\tilde{y};\tilde{y}\} = \mathscr{V}\{\tilde{y}\} = C \quad , \tag{1.2}$$

*

then it is well known that the ordinary least-squares estimator of β from the n observations $\tilde{y} = y$, with design matrix X and covariance matrix C, is given by

$$\hat{\hat{\beta}}(y) = \left(X'C^{-1}X\right)^{-1} X'C^{-1} y \quad . \tag{1.3}$$

In particular, if one makes the assumption that C is diagonal, with common terms, then (1.3) has the simpler form $\hat{\hat{\beta}} = (X'X)^{-1} X'y$, and the common error variance need not be known. Many other classical results are available based upon the normality assumption (see, e.g., [8,22,23]).

---

*We define the (possibly non-square and unsymmetric) covariance matrix,

$$\mathscr{C}\{\tilde{w};\tilde{y}\} = \mathscr{E}\{\tilde{w}\tilde{y}'\} - \mathscr{E}\{\tilde{w}\}\,\mathscr{E}\{\tilde{y}'\} \quad ,$$

for any two conformable random vectors or scalars $\tilde{w}$ and $\tilde{y}$, and write $\mathscr{C}\{\tilde{y};\tilde{y}\} = \mathscr{V}\{\tilde{y}\}$, which is usually called the covariance matrix.

2. Bayesian Multiple Regression

For a full Bayesian analysis, it is convenient to replace (1.1) by an equivalent model in which the expected values of the outputs are linear functions of the known inputs, viz.

$$\mathscr{E}\{\tilde{y}|\theta\} = X\beta(\theta) \quad . \tag{2.1}$$

Here θ denotes an unknown parameter which controls all the parameters of the conditional density, or *likelihood*, of $\tilde{y}$, given θ, denoted by $p(y|\theta)$. The *conditional covariance* of y, given θ, will be taken as an arbitrary symmetric n×n matrix

$$\mathscr{V}\{\tilde{y}|\theta\} = \sum(\theta) \quad . \tag{2.2}$$

Given the fixed, but unknown, parameters $[\beta(\theta),\sum(\theta),\ldots]$, we assume in Bayesian analysis that a *prior density*, $p(\theta)$, or what is the same thing, a joint prior density, $p(\beta,\sum,\ldots)$, is available. Then, *a priori* (i.e. prior to data), we define the first two moments of the vector of regression coefficients as

$$\mathscr{E}\beta(\tilde{\theta}) = b \quad ; \quad \mathscr{V}\beta(\tilde{\theta}) = \Delta \quad , \tag{2.3}$$

and the prior expected value of the covariance matrix as

$$\mathscr{E}\sum(\tilde{\theta}) = \mathscr{E}\mathscr{V}\{\tilde{y}|\tilde{\theta}\} = E \quad .^{*} \tag{2.4}$$

From these definitions, we can also obtain the prior first two moments of the output variables, given X. From (2.2), the mean and covariance of the conditional mean output are

$$\mathscr{E}\{\tilde{y}\} = \mathscr{E}\mathscr{E}\{\tilde{y}|\tilde{\theta}\} = Xb \quad , \tag{2.5}$$

and

---

*We use the convention that a multiple conditional expectation

$$\mathscr{E}\mathscr{E}\mathscr{E}\{f(\tilde{a},\tilde{b},\tilde{c})|\tilde{b}|\tilde{c}\}$$

means the expectation of f first with respect to $p(a|b,c)$, followed by expectation with respect to $p(b|c)$, then using $p(c)$. Arguments may be multiple, and other operators, such as $\mathscr{V}$ and $\mathscr{C}$, may be used. If the order is unimportant, and only $\mathscr{E}$ operators are used, the above is, of course, $\mathscr{E}\{f(\tilde{a},\tilde{b},\tilde{c})\}$.

$$\mathcal{V}\{\tilde{y}|\tilde{\theta}\} = D = X\Delta X' \quad . \tag{2.6}$$

From the covariance of the mean and the mean covariance, we obtain the total covariance (1.2) of the output variables prior to data as

$$\mathcal{V}\{\tilde{y}\} = C = E + D = E + X\Delta X' \quad . \tag{2.7}$$

If multinormal and related densities are used for $p(y|\theta)$ and $p(\theta)$, these are the only moments of interest.

Now, suppose an n_1-dimensional experiment is run with design matrix X_1, resulting in a vector of outputs, $\tilde{y} = y_1$; we denote this by (n_1, X_1, y_1). Using the likelihood $p(y_1|\theta) = p(y_1|\theta, X_1)$, and the prior on the parameters, $p(\theta)$, we obtain the *posterior* (to the data) *density* $p(\theta|y_1) = p(\theta|y_1, X_1)$ in the usual way:

$$p(\theta|y_1) = \frac{p(y_1|\theta)p(\theta)}{\int p(y_1|\Phi)p(\Phi)d\Phi} \quad , \tag{2.8}$$

where, for convenience, we suppress the known design matrix, X_1.

From (2.8), the updated estimates of the parameters $\beta(\tilde{\theta})$, $\sum(\tilde{\theta}),\ldots,$ are, in principle, available. For example, the expected value of the vector of regression coefficients posterior to the data is

$$\mathcal{E}\{\beta(\tilde{\theta})|y_1\} = \int \beta(\theta)p(\theta|y_1)d\theta \quad , \tag{2.9}$$

and the *predictive density* for a future experiment (n_2, X_2, y_2), with the same parameters, but independent outputs, is

$$p(y_2|y_1) = p(y_2|y_1, X_1, X_2) = \int p(y_2|\theta, X_2)p(\theta|y_1)d\theta \quad . \tag{2.10}$$

Because of the difficulty of carrying out (2.8)-(2.10) for arbitrary priors and likelihoods, most of the Bayesian regression literature makes the following additional assumptions:

(1) The likelihood, $p(y|\theta) = p(y|\theta,X)$, is multinormal for any experiment (n,X,y)--thus only the parameters $\tilde{\beta} = \beta(\tilde{\theta})$ and $\tilde{\Sigma} = \Sigma(\tilde{\theta})$ are involved, and (2.8) can be restated in terms of $p(\beta,\Sigma)$;

(2) Either the Ando-Kaufmann [1] Normal-Wishart natural-conjugate prior $p(\beta,\Sigma)$ is used to simplify the updating in (2.8);

(3) Or, $\tilde{\beta}$ and $\tilde{\Sigma}$ are assumed independent, $p(\beta,\Sigma) = p(\beta)p(\Sigma)$, and simple marginal densities are chosen, typically multinormal or non-informative (diffuse) for $\tilde{\beta}$, and inverse Wishart or non-informative for $\tilde{\Sigma}$.

There are difficulties with all of these assumptions. For example, the Ando-Kaufmann prior is well known to be "thin"; that is, not all possible hyperparameters in $p(\beta,\Sigma)$ can be specified independently. And analysts are divided over the use of non-informative priors, although in some cases they follow from invariance or limiting arguments ([32], p. 226).

Also, computations made under these assumptions are distinctly untidy, involving much completion of the square, matrix manipulation, and multidimensional integration, particularly if the full posterior parameter density, $p(\beta,\Sigma|y_1)$, and its marginals are desired, or if the predictive density (2.10) is sought [21,30,32]. The only non-trivial relaxations of the normality assumption of which we are aware are the numerical trials of Box and Tiao ([3], Chapter 3) with the exponential power distribution.

In the sequel, we propose to follow a more modest course, by concentrating on (2.9) and the related problem of predicting the mean outcome of a future experiment, by using the linearized ideas of credibility theory. This almost distribution-free approach will greatly simplify the resulting formulae, and will provide an intuitively appealing bridge between classical and Bayesian regression techniques. And we shall see that in many cases of practical interest, the linearized credibility formulae are also exact Bayesian.

First we review the basic concepts of credibility theory.

3. Credibility Theory

Credibility theory is essentially linear least-squares applied to conditional distributions. Suppose that a p-dimensional random vector, $\tilde{w}$, is to be forecast from a single sample of an r-dimensional random vector, $\tilde{y} = y$, in the sense of finding a p-dimensional vector forecast function, $f(y)$, which minimizes the sum of the expected squared errors for each component

$$H = \int\int \sum_{i=1}^{p} [w_i - f_i(y)]^2 \, dP(w,y) = tr\mathscr{E}\{[\tilde{w} - f(\tilde{y})][\tilde{w} - f(\tilde{y})]'\} \quad . \tag{3.1}$$

It is known that the integrable functions f_i^0 which minimize (3.1) at value H^0 form the conditional mean vector,

$$f^0(y) = \mathscr{E}\{\tilde{w}|y\} \quad . \tag{3.2}$$

In many cases the exact conditional mean is difficult to calculate, and an approximate forecast vector, f, is acceptable. By completing the square, we find

$$\begin{aligned} H &= H^0 + \int \sum_{i=1}^{p} [f_i^0(y) - f_i(y)]^2 \, dP(y) \quad , \\ H^0 &= tr\,\mathscr{E}\mathscr{V}\{\tilde{w}|\tilde{y}\} \quad , \end{aligned} \tag{3.3}$$

so that any f can also be evaluated in terms of its fit to the conditional mean $f^0(y)$.

A convenient choice of an approximate forecast vector is a linear function of the observables,

$$f_i(y) = z_{i0} + \sum_{j=1}^{r} z_{ij} y_j \quad , \qquad (i = 1, \ldots p) \quad , \tag{3.4}$$

where the $p(r+1)$ coefficients $\{z_{ij}\}$, henceforth called *credibility coefficients*, are adjusted so as to minimize (3.1) or (3.3). It is well known that the optimal values of these coefficients are then given by rp *normal equations* of the form

$$\sum_{j=1}^{r} z_{ij} \mathscr{C}\{\tilde{y}_j; \tilde{y}_k\} = \mathscr{C}\{\tilde{w}_i; \tilde{y}_k\} \quad , \qquad \begin{matrix} (i = 1, \ldots p) \\ (k = 1, \ldots r) \end{matrix} \tag{3.5}$$

with the $\{z_{iO}\}$ determined so as to make the forecast (3.4) unbiased:

$$z_{iO} = \mathscr{E}\{\tilde{w}_i\} - \sum_{j=1}^{r} z_{ij}\,\mathscr{E}\{\tilde{y}_j\};\quad \mathscr{E}\{f_i(\tilde{y})\} = \mathscr{E}\{\tilde{w}_i\} \quad , \quad (i = 1, \ldots p) \quad . \qquad (3.6)$$

Let z_O be the p-vector $[z_{iO}]'$, and Z the p×r matrix $[z_{ij} | j \neq O]$; then the optimal conditions (3.5)(3.6) can be written as

$$Z\mathscr{V}\{\tilde{y}\} = \mathscr{C}\{\tilde{w};\tilde{y}\} \quad , \qquad (3.7)$$

and

$$z_O = \mathscr{E}\{\tilde{w}\} - Z\mathscr{E}\{\tilde{y}\} \quad , \qquad (3.8)$$

so that the optimal linear forecast (3.4) is

$$f(y) = \mathscr{E}\{\tilde{w}\} + Z[y - \mathscr{E}\{\tilde{y}\}] \quad , \qquad (3.9)$$

and all attention can be focussed on finding the credibility matrix, Z, from (3.7). The minimal value of H is then easily shown to be

$$H = \mathrm{tr}\left[\mathscr{V}\{\tilde{w}\} - Z\mathscr{C}\{\tilde{y};\tilde{w}\}\right] \geq H^O \quad . \qquad (3.10)$$

Notice that each component in (3.1) is, in fact, minimized independently; we use matrix notation only for convenience.

In Bayesian problems, the joint distribution of $\tilde{w}$ and y is parametrized by a parameter θ which is not known. Therefore the optimal Z must be determined *a priori*, using measure $P(w,y) = \mathscr{E}P(w,y|\tilde{\theta})$. Thus, the covariances in (3.7) will, in general, consist of two terms similar to (2.7). One also looks for special forms of $\mathscr{V}\{\tilde{y}\}$ which will simplify the computation of Z in (3.7) [16].

In the insurance models which gave rise to credibility theory, there is an underlying sequence of p-dimensional random vectors $\{\tilde{x}_1, \tilde{x}_2, \ldots \tilde{x}_t, \tilde{x}_{t+1}, \ldots\}$, which are independent and identically distributed, given a fixed, but unknown,

"risk parameter," θ. The problem is to predict $\mathcal{E}\{\tilde{x}_{t+1}|x_1,x_2,\ldots x_t\}$, called the "experience-rated fair premium". Using the above analysis, it is easy to show that the optimal linearized approximation to the conditional mean is

$$\mathcal{E}\{\tilde{x}_{t+1}|x_1,x_2,\ldots x_t\} \approx f(x_1,x_2,\ldots x_t) = (I_p - Z_x)\mathcal{E}\{\tilde{x}\} + Z_x\left[\frac{1}{t}\sum_{u=1}^{t} x_u\right], \tag{3.11}$$

where I_p is the p×p unit matrix, and Z_x is the p×p optimal credibility matrix, given by

$$Z_x(E_x + tD_x) = tD_x \quad , \tag{3.12}$$

where E_x and D_x are the p×p matrix components of the covariance of a typical $\tilde{x}$, defined in a manner similar to (2.4) and (2.6) [13].

The original credibility formula was developed heuristically by American actuaries in the '20s for a one-dimensional version of (3.11), in which Z_x gives the weight, or "credibility," to be attached to the "experience" sample mean, $(\Sigma x_u/t)$, as opposed to the "manual fair premium" $\mathcal{E}\{\tilde{x}\}$. In the one-dimensional case, $0 \leq Z_x \leq 1$, and approaches unity as the "weight of evidence", t, becomes large. In the general (but nondegenerate) model, Z_x consists of p^2 rational functions of t, not restricted to [0,1]; however, $Z_x \to I_p$ as $t \to \infty$, showing that ultimately the sample mean of the ith component is "fully credible" for predicting the ith component of the next observation.

Although credibility theory was originally developed as an approximation theory for mean forecasts, it can also be used as an approximation theory for higher moments, or even for distributions [4,5,11].

Moreover, and perhaps more importantly, it also turns out to be an exact theory for forecasting the mean, when the likelihood is a member of the exponential family in which the sample mean is a sufficient statistic, and when a natural conjugate prior is chosen. For further details, see [12,13,14].

4. Credibility Applied To Regression

We now apply the above theory to three related Bayesian estimation problems, assuming that data from an (n_1, X_1, y_1) experiment is available:

(1) the estimation of the mean regression parameters posterior to the data;

(2) the prediction of the mean response in a future experiment (n_2, X_2, y_2);

(3) the estimation of the mean error variables in (1.1).

We shall show, with minor exceptions, that the three credibility estimates are equivalent, and related to the classical estimator (1.3).

4.1 Estimation of Regression Parameters

Suppose we wish to estimate $\mathscr{E}\{\beta(\tilde{\theta})\,|\,y_1\}$ with credibility theory (X_1 is still fixed and known). Then in Section 3 we take $\tilde{w} = \beta(\tilde{\theta})$, $k = r$, and $\tilde{y} = \tilde{y}_1$, giving $\mathscr{E}\{\tilde{w}\} = b$, $\mathscr{E}\{\tilde{y}\} = X_1 b$,

$$\mathscr{C}\{\tilde{w};\tilde{y}\} = \mathscr{C}\{\beta(\tilde{\theta});\mathscr{E}\{\tilde{y}_1\,|\,\tilde{\theta}\}\} = \Delta X_1' \quad ,$$

and, from (2.7),

$$\mathscr{V}\{\tilde{y}\} = C_{11} = E_{11} + X_1 \Delta X_1' \quad ,$$

where $E_{11} = \mathscr{E}\Sigma_{11}(\tilde{\theta})$ is the $n_1 \times n_1$ matrix of expected covariances of y_1 during the experiment.

From (3.7), the $k \times n_1$ credibility matrix

$$Z_\beta = \Delta X_1' C_{11}^{-1} = \Delta X_1' (E_{11} + X_1 \Delta X_1')^{-1} \qquad (4.1)$$

gives a linear, unbiased estimate of the posterior parameter vector

$$\mathscr{E}\{\beta(\tilde{\theta})\,|\,y_1, X_1\} \approx f_\beta(y_1, X_1) = (I_k - Z_\beta X_1)b + Z_\beta y_1 \quad . \qquad (4.2)$$

Notice that no assumptions have been made about the distributions $p(y|\theta)$ and $p(\theta)$ (except for the existence of the

indicated moments), nor about the independence of the components of $\tilde{y}_1$, given θ. However, E_{11}^{-1} must exist for the inverse in (4.1) to be well defined, if no special assumptions are made about X_1 (see Section 4.3).

4.2 Prediction of Mean Response in Future Experiments

Now suppose we have in mind a well-defined future experiment (n_2, X_2, y_2), and the problem is to estimate $\mathscr{E}\{\tilde{y}_2|y_1\} = \mathscr{E}\{\tilde{y}_2|y_1, X_1, X_2\}$ by credibility theory. There are two possible cases, depending on whether

$$\Sigma_{21}(\tilde{\theta}) = \mathscr{C}\{\tilde{y}_2;\tilde{y}_1|\theta\} \quad ; \quad E_{21} = \mathscr{E}\{\Sigma_{21}(\tilde{\theta})\} \quad ;$$

are zero or not, i.e., whether knowledge of the parameter decouples the results of past and future experiments or not.

4.2.1 No Covariance Between Experiments

In most classical regression models, there is no covariance between past and future observations, given θ, either by assumption, or because there is a sufficient interval between the two experiments, even if, say, the error process has serial correlation.

For an exact Bayesian analysis, we have from (2.1) and (2.9):

$$\mathscr{E}\{\tilde{y}_2|y_1, X_1, X_2\} = X_2\mathscr{E}\{\beta(\tilde{\theta})|y_1, X_1\} \quad , \tag{4.3}$$

which shows the close relation between the two problems.

Similarly, because of the linearity of a credibility forecast, it follows that

$$\begin{aligned}\mathscr{E}\{\tilde{y}_2|y_1, X_1, X_2\} \approx f_{y_2}(y_1, X_1, X_2) &= X_2 f_\beta(y_1, X_1) \\ &= (X_2 - Z_{y_2}X_1)b + Z_{y_2}y_1 \quad , \end{aligned} \tag{4.4}$$

where Z_{y_2} is the $n_2 \times n_1$ credibility matrix

$$Z_{y_2} = X_2\Delta X_1'(E_{11} + X_1\Delta X_1')^{-1} = X_2 Z_\beta \quad . \tag{4.5}$$

In other words, when there is no covariance between experiments, estimation of the regression coefficients by credibility is equivalent to estimation of future response.

4.2.2 Covariance Between Experiments

In the general case in which $\Sigma_{21}(\theta) \neq 0$, infrequently considered in the literature, the complete Bayesian analysis is more complicated, and one needs to replace the assumption $\mathscr{E}\{\tilde{y}_2|X_2;\theta\} = X_2\beta(\theta)$ by an equivalent assumption about $\mathscr{E}\{\tilde{y}_2|y_1,X_1,X_2,\theta\}$. This could be of arbitrary form, but if it is to be in agreement with the classical multinormal results, then we must choose the usual *regression of* y_2 *on* y_1 (see, e.g. [23]):

$$\mathscr{E}\{\tilde{y}_2|y_1,X_1,X_2,\theta\} = X_2\beta(\theta) + \Sigma_{21}(\theta)\,\Sigma_{11}^{-1}(\theta)[y_1 - X_1\beta(\theta)] \quad . \quad (4.6)$$

In an exact updating through (2.8), difficulty would arise from the possible covariance of the terms $\Sigma_{21}(\theta)$ and $\Sigma_{11}^{-1}(\theta)$ with each other, and with $\beta(\theta)$. However, if these terms have small covariances compared with those of $\beta(\theta)$, then one could with small error replace these terms by their expected values, and use the approximation

$$\mathscr{E}\{\tilde{y}_2|y_1,X_1,X_2,\beta(\theta)\} \approx X_2\beta(\theta) + E_{21}E_{11}^{-1}[y_1 - X_1\beta(\theta)] \quad (4.7)$$

to give an exact Bayesian updating:

$$\mathscr{E}\{\tilde{y}_2|y_1,X_1,X_2\} \approx X_2\mathscr{E}\{\beta(\tilde{\theta})|y_1\} + E_{21}E_{11}^{-1}[y_1 - X_1\mathscr{E}\{\beta(\tilde{\theta})|y_1\}] \quad . \quad (4.8)$$

In the credibility approximation, the formula in Section 4.2.1 is replaced by

$$\mathscr{C}\{\tilde{w};\tilde{y}\} = X_2\Delta X_1' + E_{21} \quad , \quad (4.9)$$

so that the new credibility matrix is

$$Z_{y_2} = (X_2\Delta X_1' + E_{21})(E_{11} + X_1\Delta X_1')^{-1} \quad , \quad (4.10)$$

and, after some algebra, we find

$$\mathcal{E}\{\tilde{y}_2|y_1,X_1,X_2\} \approx f_{y_2}(y_1,X_1,X_2) \equiv X_2 f_\beta(y_1,X_1) + E_{21}E_{11}^{-1}[y_1 - X_1 f_\beta(y_1,X_1)], \tag{4.11}$$

which is of the same form as (4.8). So, to the degree to which (4.7) may replace (4.6), we again have a simple relation between credibility estimates for the parameters and forecasts for future observations.

4.3 Relationship to Classical Regression Estimation

In classical regression, emphasis is placed upon having sufficient observations to fully identify all of the regression parameters, i.e., $n_1 \geq k$, and X_1 has full rank k; the necessity for this can be seen from the classical estimator (1.3).

On the other hand, in the Bayesian credibility model, it can be seen from (4.1)-(4.2) that the finiteness of b, E_{11}^{-1}, and Δ is sufficient to guarantee the existence of an estimator for $\tilde{\beta}$; one sample will revise the prior estimate of b, even if X_1 does not have full rank! In fact, if n_1 is small, the calculation of $(E_{11} + X_1\Delta X_1')^{-1}$ is particularly simple.

However, to relate our results to classical theory, we shall henceforth assume that $n_1 \geq k$, and $\text{rank}(X_1) = k$, and use the following result which Bodewig ([2] pp. 39, 218) attributes to H. Hemes, and which is also given by Tocher [29] (see also Lindley and Smith [19], pp. 6 and 34 for two later attributes).

Theorem. If α and β are $n \times k$ matrices, then

$$(I_n + \alpha\beta')^{-1} = I_n - \alpha(I_k + \beta'\alpha)^{-1}\beta' \quad , \tag{4.12}$$

whenever either of the indicated inverses exists.

The fact that the determinants of the two terms in parenthesis are identical shows that the existence of one inverse implies the existence of the other.

If we apply this to C_{11}^{-1}, with $\alpha = X_1$ and $\beta' = \Delta X_1' E_{11}^{-1}$, we get

$$\begin{aligned} C_{11}^{-1} &= (E_{11} + X_1\Delta X_1')^{-1} \\ &= E_{11}^{-1}[I_n - X_1(I_k + \Delta X_1' E_{11}^{-1} X_1)^{-1}\Delta X_1' E_{11}^{-1}] \quad . \end{aligned} \tag{4.13}$$

Defining the two $k \times k$ matrices

$$\varepsilon_1^{-1} = X_1' E_{11}^{-1} X_1 \quad ; \tag{4.14}$$

$$z_1 = (I_k + \varepsilon_1 \Delta^{-1})^{-1} = \Delta(\Delta + \varepsilon_1)^{-1} = (\varepsilon_1^{-1} + \Delta^{-1})^{-1} \varepsilon_1^{-1} ; \tag{4.15}$$

we obtain finally

$$Z_\beta = z_1 \varepsilon_1 X_1' E_{11}^{-1} \quad ; \qquad Z_\beta X_1 = z_1 \quad ; \tag{4.16}$$

and (4.2) and (4.5) become

$$f_\beta(y_1, X_1) = (I_k - z_1) b + z_1 \hat{\beta}_1(y_1) \quad ; \tag{4.17}$$

$$f_{y_2}(y_1, X_1, X_2) = X_2 [(I_k - z_1) b + z_1 \hat{\beta}_1(y_1)] \quad ; \tag{4.18}$$

with a k-dimensional vector estimator for $\hat{\beta}$ of

$$\hat{\beta}_1(y_1) = (X_1' E_{11}^{-1} X_1)^{-1} X_1' E_{11}^{-1} y_1 \quad . \tag{4.19}$$

This rearrangement requires $\text{rank}(\varepsilon_1^{-1}) = k$.

(4.17) is, from an aesthetic viewpoint, extremely satisfying, for it shows the familiar credibility mixing between the prior mean parameter vector, b, and a sample statistic, $\hat{\beta}(y_1)$, in a manner similar to the multidimensional credibility formula (3.11), and extensions of it to other sample statistics [12][13]. Only a small credibility matrix, z_1, need be calculated from (4.15), and its size depends only on the number of parameters to be estimated, not the number of data points. Of course, one must calculate E_{11}^{-1}, but this is needed in any regression problem, and is often assumed to be of diagonal form. There is an obvious parallel between (4.15) and (3.12).

There remains to explain the relation between the estimator $\hat{\beta}_1(y_1)$ in (4.19), and the classical estimator $\hat{\hat{\beta}}_1(y_1)$ in (1.3), for, as we know, the latter should be used with the total covariance $C_{11} = E_{11} + X_1 \Delta X_1'$. However, a simple calculation will show that the second term is annihilated in the

least-squares form, so that

$$\hat{\hat{\beta}}_1(y_1) \equiv \hat{\beta}_1(y_1) \quad , \tag{4.20}$$

and it is a matter of indifference how the estimator is calculated.

4.4 Estimation of Error Variables

After a regression model has been calibrated, it is often useful to verify the assumptions of the model by examining the residual vector, $y_1 - X_1 f_\beta(y_1, X_1)$.

One can also think of estimating the true value of the error variables, u_1, in (1.1) by using Bayesian analysis [33]. Using the credibility approach, we first find $\mathscr{E}\{\tilde{u}_1\} = 0$, $\mathscr{V}\{\tilde{u}_1\} = \mathscr{C}\{\tilde{u}_1;\tilde{y}_1\} = E_{11}$, and then find the mean estimate ,

$$\begin{aligned} \mathscr{E}\{\tilde{u}_1 | y_1, X_1\} \approx f_{u_1}(y_1, X_1) &= (I_{n_1} - X_1 Z_\beta)(y_1 - X_1 b) \\ &= y_1 - X_1 f_\beta(y_1, X_1) \quad , \end{aligned} \tag{4.21}$$

which is exactly the vector of residuals! This might have been expected from first principles.

Perhaps it is worth pointing out that [6, Appendix 3]

$$\mathscr{C}\{\tilde{u}_1; f_{u_1}(\tilde{y}_1, X_1)\} = 0 \quad . \tag{4.22}$$

5. Estimation Error Covariances--Limiting Cases

It is of interest to compute the improvement in estimation to be expected from the credibility formulae.

For the regression parameters, let the estimation error covariance matrix be

$$\Phi_\beta(X_1) = \mathscr{E}\{[\beta(\tilde{\theta}) - f_\beta(\tilde{y}_1, X_1)][\beta(\tilde{\theta}) - f_\beta(\tilde{y}_1, X_1)]'\}$$

$$= \mathscr{V}\{\beta(\tilde{\theta}) - f_\beta(\tilde{y}_1, X_1)\} \quad , \tag{5.1}$$

because the estimator is unbiased, a priori.

By elementary calculations based on Sections 3.1 and 4, we find that the minimal "preposterior" value is the analog of the term in square brackets in (3.10):

$$\Phi_\beta(X_1) = (I_k - z_1)\Delta = z_1\varepsilon_1 \quad . \tag{5.2}$$

Remember that only the diagonal terms of Φ are (independently) minimized in using (3.1), $H = tr\Phi$.

For the prediction of mean future response, we find in the no-covariance case of Section 4.2.1:

$$\Phi_{y_2}(X_1, X_2) = \mathscr{V}\{\tilde{y}_2 - f_{y_2}(\tilde{y}_1, X_1, X_2)\}$$

$$= E_{22} + X_2(I_k - z_1)\Delta X_2' = E_{22} + X_2 z_1 \varepsilon_1 X_2' \quad . \tag{5.3}$$

The result with covariance between experiments is similar, with additional terms involving E_{21}.

The preposterior estimate of the covariance matrix of the residual vector (4.21) is

$$\Phi_{u_1}(X_1) = X_1 Z_\beta E_{11} = X_1 \Phi_\beta(X_1) X_1' \quad . \tag{5.4}$$

Without an initial experiment, the value of z_1 would be zero, and from (4.17)(4.18)(4.21) we would have to use the means, b, $X_2 b$ and y_1, as predictors, and (5.2)(5.3)(5.4) would be equal to the appropriate total prior covariance matrices,

Δ, $E_{22} + X_2 \Delta X_2'$, and 0, respectively.

Similarly, if the first experiment is performed under poor observational conditions, then the diagonal elements of E_{11} will be much larger than those of $X_1 \Delta X_1'$. We see directly that z_1 would be zero, and there would be a vote of "no confidence" in the estimator $\hat{\beta}_1(y_1)$, and b, $X_2 b$, and y_1 would again be the minimum-variance predictors for $\beta(\theta)$, y_2, and u_1, respectively.

However, conversely, if the diagonal elements of Δ are very large compared to those of ε_1, this means that our prior knowledge is very imprecise compared to the error conditions of the experiment; $\Delta^{-1} \to 0$ is the credibility equivalent of the "diffuse prior" assumptions often made in Bayesian analysis. In this case, we see that $z_1 \to 1$; "full credibility" is attached to the classical estimator $\hat{\beta}_1(y_1)$, and the prior mean, b, is given zero weight. There remain only the irreducible error covariances ε_1 in estimating $\beta(\theta)$, $E_{22} + X_2 \varepsilon_1 X_2'$ in predicting y_2, and $X_1 \varepsilon_1 X_1'$ in estimating u_1.

Also, if we consider experiments with increasing n_1, then, under certain natural conditions, such as:

(1) The elements of E_{11} are bounded, for all n_1;

(2) The design matrix, X_1, "fills out" a finite range of the x-axis in a stable manner, as n_1 increases;

it is easy to show that the elements of ε_1 in (4.14) are bounded by a function which diminishes as n_1^{-1}, that is, z_1 approaches I_k as n_1 increases (see, e.g., [18]). In practical terms, this means that an increasing number of initial sample points can reduce the preposterior covariance in estimating the regression parameter (5.2) as close to zero as desired; however, there will always be an irreducible covariance E_{22} in making forecasts (5.3). The covariance matrix $\Phi_u(X_1)$ in (5.4) continues to grow in dimension, and depends in a complicated manner upon the actual structure of X_1.

6. Random Design Matrices

In many applications, X_1 and/or X_2 must be considered as random, either as a result of an uncontrollable input, because the effective input cannot be precisely observed, or because of deliberate randomization. There are many special cases in the literature, (see, e.g.,[7,32]); we shall derive general credibility results, and indicate only a few of the possible specializations. Special attention must be paid to whether X_1, X_2, or both are random variables, so throughout this section we shall indicate the status of all inputs and outputs explicitly. We start with two simpler cases.

6.1 X_2 Random and Independent of Fixed Initial Experiment

If the future design matrix X_2 is random, but independent of the fixed initial experiment (n_1, X_1, y_1), then the problem of estimating the regression parameters is unchanged from Section 4.1.

However, to predict the mean response of the second experiment, we must now calculate a credibility approximation to $\mathscr{E}\{\tilde{y}_2|y_1,X_1\} = \mathscr{E}\mathscr{E}\{\tilde{y}_2|y_1,X_1,\tilde{X}_2\}$. Assuming, for simplicity, unobservationally unrelated experiments, $\Sigma_{21}(\theta) = 0$, we have from (2.1) and Section 4.2.1.,

$$\mathscr{E}\{\tilde{y}_2\} = \mathscr{E}[\mathscr{E}\{\tilde{X}_2|\tilde{\theta}\}\cdot\beta(\tilde{\theta})] \quad , \tag{6.1}$$

and

$$\mathscr{C}\{\tilde{y}_2;\tilde{y}_1|X_1\} = \mathscr{C}\{\mathscr{E}\{\tilde{X}_2|\tilde{\theta}\}\cdot\beta(\tilde{\theta});X_1\beta(\tilde{\theta})\} \quad . \tag{6.2}$$

Since $\mathscr{V}\{\tilde{y}_1|X_1\} = E_{11} + X_1\Delta X_1'$ and $\mathscr{E}\{\tilde{y}_1|X_1\} = X_1 b$ still, the only effect in this case has been to modify the first term, $X_2\Delta X_1'$, in the definition of z_{y_2} in (4.5) to the form in (6.2) and to change the z_0 term in (4.4).

An important special case is:

Assumption I. Any random X is statistically independent of θ. (6.3)

In this case, we see directly that $\mathscr{E}\{\tilde{y}_2\} = \mathscr{E}\{\tilde{x}_2\}b$ and $\mathscr{C}\{\tilde{y}_2;\tilde{y}_1|X_1\} = \mathscr{E}\{\tilde{X}_2\}\Delta X_1'$, that is, all the results of Section

4.2.1 apply with X_2 replaced by its expected value!

6.2 Estimation of Regression Parameters when X_1 is Random

If X_1 is random, then to estimate $\beta(\theta)$ we must use the joint density $p(y_1, X_1|\theta)$ and generalize (4.2). For the mean outcome of the initial experiment,

$$\mathscr{E}\{\tilde{y}_1\} = \mathscr{E}\{\mathscr{E}\{\tilde{X}_1|\tilde{\theta}\}\cdot\beta(\tilde{\theta})\} \quad , \tag{6.4}$$

but the covariance of y_1 now has three terms:

$$\mathscr{V}\{y_1\} = \mathscr{E}\{\Sigma_{11}(\tilde{X}_1,\tilde{\theta})\} + \mathscr{V}\{\mathscr{E}\{\tilde{X}_1|\tilde{\theta}\}\cdot\beta(\tilde{\theta})\} + \mathscr{E}\mathscr{V}\{\tilde{X}_1\beta(\tilde{\theta})|\tilde{\theta}\} \quad , \tag{6.5}$$

where

$$\Sigma_{11}(X_1,\theta) = \mathscr{V}\{\tilde{y}_1|X_1,\theta\} \tag{6.6}$$

shows explicitly the possible dependence of the conditional observational covariance both on the design X_1 and on θ. (For consistency, we shall assume in the next section that neither (6.4) nor (6.6) can, however, depend upon the future values (y_2, X_2).)

Since $\beta(\theta)$ is constant, given θ, there is still only one term in

$$\mathscr{C}\{\beta(\tilde{\theta});\tilde{y}_1\} = \mathscr{C}\{\beta(\tilde{\theta});\mathscr{E}\{\tilde{X}_1|\tilde{\theta}\}\cdot\beta(\tilde{\theta})\} \quad . \tag{6.7}$$

This form and the first two terms in (6.5) are easily seen to be the generalizations of $\Delta X_1'$ and $E_{11} + X_1\Delta X_1'$, respectively, as used in Section 4.1.

However, the last term in (6.5) is new, call it U. It has components

$$U_{tu} = \mathscr{E}\left\{\sum_{i=1}^{k}\sum_{j=1}^{k}\beta_i(\tilde{\theta})\beta_j(\tilde{\theta})\mathscr{C}\{\tilde{x}_{ti};\tilde{x}_{uj}|\tilde{\theta}\}\right\} \quad ,$$

$$(t,u = 1,2,\ldots k) \quad , \tag{6.8}$$

and thus contains information about the conditional covariances

between independent variables.

In many models, such as "errors-in-the-variables," or "target inputs" [7], successive inputs are independent, or have independent errors around fixed means, expressable as:

Assumption II. Rows of any random X are statistically independent. (6.9)

In this case, it follows that U is diagonal. Additionally, we point out that in many regression designs, the first column of X_1 is non-random (consisting entirely of 1's), so that the summations in (6.8) would begin with i = 2 and j = 2.

If Assumption I is taken also to apply to $\tilde{X}_1$,

$$\begin{aligned} &\mathscr{E}\{\tilde{y}_1\} = \mathscr{E}\{\tilde{X}_1\}b \quad ; \\ &\mathscr{V}\{\tilde{y}_1\} = \mathscr{E}\{\Sigma_{11}(\tilde{X}_1,\tilde{\theta})\} + \mathscr{E}\{\tilde{X}_1\}\Delta\mathscr{E}\{\tilde{X}_1'\} + U \quad ; \\ &\mathscr{C}\{\beta(\tilde{\theta});\tilde{y}_1\} = \Delta\mathscr{E}\{X_1'\} \quad ; \end{aligned} \tag{6.10}$$

and the main effect on the credibility estimate (4.1), apart from replacing X_1 by its mean value, and defining a more general average covariance E_{11}, is to add a diagonal matrix U to the covariance of $\tilde{y}_1$, with terms

$$U_{tt} = \sum_{i=1}^{k}\sum_{j=1}^{k}(\Delta_{ij} + b_i b_j)\mathscr{C}\{\tilde{x}_{ti};\tilde{x}_{tj}\} \quad ,$$

$$(t = 1,2,\ldots k) \quad . \tag{6.11}$$

This will change Z_β in an obvious manner, and we see that the estimator to be used in (4.17) becomes

$$\hat{\beta}(y_1) = \left[\mathscr{E}\{X_1'\}(E_{11}+U)^{-1}\mathscr{E}\{\tilde{X}_1\}\right]^{-1}\mathscr{E}\{\tilde{X}_1'\}(E_{11}+U)^{-1}y_1 \quad , \tag{6.12}$$

with the new interpretation of E_{11} from (6.10), and a new

$$\varepsilon_1^{-1} = \mathscr{E}\{X_1'\}(E_{11}+U)^{-1}\mathscr{E}\{\tilde{X}_1\} \tag{6.13}$$

used to define z_1 in (4.15).

6.3 General Case

In the general case when all inputs and outputs are random, we must work with the joint density $p(y_1,X_1,y_2,X_2|\theta)$, and be extremely careful about the assumptions of dependence and independence which are appropriate to the model under consideration. Different models may lead to different conditional decompositions of this joint density.

Usually the regression parameters are estimated after the initial experiment, so that the results of Section 6.2 apply. If both experiments are performed, then the total data may be pooled, and the same results apply with obvious modification (see Section 7).

Therefore the central problem of interest in credibility theory will be to predict $\mathscr{E}\{\tilde{y}_2|\tilde{y}_1\}$, for which we need: $\mathscr{E}\{\tilde{y}_1\}$, $\mathscr{E}\{\tilde{y}_2\}$, $\mathscr{V}\{\tilde{y}_1\}$ and $\mathscr{C}\{\tilde{y}_2;\tilde{y}_1\}$. (6.4) and (6.5) still apply because the data-gathering experiment is prior to the one for which the prediction is made. However, to compute $\mathscr{E}\{\tilde{y}_2\}$, we need an assumption such as (4.7) to specify a form for $\mathscr{E}\{\tilde{y}_2|y_1,X_1,X_2,\theta\}$. Given this, we then uncondition in any convenient way, say

$$\mathscr{E}\{\tilde{y}_2\} = \mathscr{E}\mathscr{E}\mathscr{E}\mathscr{E}\mathscr{E}\{\tilde{y}_2|\tilde{X}_2|\tilde{y}_1|\tilde{X}_1|\tilde{\theta}\} \quad , \tag{6.14}$$

using any other simplifications, such as Assumption I, which apply. Further reduction will need a careful analysis of the experimental conditions; for example

> Assumptions III(a) (b) or (c). The choice of the future design, $\tilde{X}_2$, given θ, depends only on (a) the past input, X_1; or (b) the past output, y_1; or (c) on both (X_1,y_1); (6.15)

III(a) might obtain if (X_1,X_2) were part of the same predetermined experimental design, or if errors in the independent variables were serially correlated; III(b) might be correct if the future input values depended upon the previous outputs, or perhaps on some estimator of $\beta(\theta)$, such as (4.2), as generalized in Section 6.2.

For the RHS of (3.7), repeated application of the principle of conditional covariance leads to

$$\begin{aligned}\mathscr{C}\{\tilde{y}_2;\tilde{y}_1\} = {} & \mathscr{E}\mathscr{E}\mathscr{E}\{\Sigma_{21}(\tilde{X}_2;\tilde{X}_1;\tilde{\theta})\,|\,\tilde{X}_1\,|\,\tilde{\theta}\} \\ & + \mathscr{E}\mathscr{C}\{\mathscr{E}\mathscr{E}\{\tilde{y}_2\,|\,\tilde{X}_2\,|\,\tilde{X}_1,\tilde{\theta}\};\tilde{X}_1\beta(\tilde{\theta})\,|\,\tilde{\theta}\} \qquad (6.16) \\ & + \mathscr{C}\{\mathscr{E}\mathscr{E}\mathscr{E}\{\tilde{y}_2\,|\,\tilde{X}_2\,|\,\tilde{X}_1\,|\,\tilde{\theta}\};\mathscr{E}\{\tilde{X}_1\,|\,\tilde{\theta}\}\cdot\beta(\tilde{\theta})\}\ ,\end{aligned}$$

where the arguments of $\Sigma_{21}(X_2,X_1,\theta)$ show that the covariance of observational errors between $\tilde{y}_2$ and $\tilde{y}_1$ can now depend upon both inputs; one possible term in (6.16) is missing because we still assume $\mathscr{E}\{\tilde{y}_1\,|\,X_2,X_1,\theta\} = X_1\beta(\theta)$. Further simplification depends upon using forms such as (4.7), and clarifying the experimental relationships between $\tilde{\theta}$, $\tilde{X}_1$, and $\tilde{X}_2$.

7. Prior Information and Prior Experiments

The distinction between prior information, in the usual Bayesian sense, and the information obtained as the result of a prior experiment is not clear-cut. Suppose we have given prior information (b,Δ) about $\beta(\theta)$, and the matrix of observation error covariances E for any (n,X). A first experiment (n_1,X_1,y_1) then provides a further estimate of $\beta(\theta)$, which supplements our knowledge prior to the performance of a second experiment (n_2,X_2,y_2); thus, there is total prior information $(b,\Delta;E_{11};n_1,X_1,y_1)$ as input to the second stage. On the other hand, we know that the estimation of $\beta(\theta)$ after two experiments can be regarded as a combined single experiment, and it is interesting to examine further the relationship between these two viewpoints.

To estimate $\mathscr{E}\{\beta(\tilde{\theta})\,|\,y_1,X_1;y_2,X_2\}$, we form the enlarged versions of (2.1)(2.2):

$$\mathscr{E}\left\{\begin{bmatrix}\tilde{y}_1\\ \tilde{y}_2\end{bmatrix}\,\middle|\,\theta\right\} = \begin{bmatrix}X_1\\ X_2\end{bmatrix}\beta(\theta) \quad ; \tag{7.1}$$

$$\mathscr{V}\left\{\begin{bmatrix}\tilde{y}_1\\ \tilde{y}_2\end{bmatrix}\,\middle|\,\theta\right\} = \begin{bmatrix}\Sigma_{11}(\theta) & 0\\ 0 & \Sigma_{22}(\theta)\end{bmatrix} \quad ; \tag{7.2}$$

where we have assumed the two experiments are observationally independent, and the design matrices are fixed. Then, following the analysis of Section 4.1, we find an enlarged Z_β-type $k \times (n_1 + n_2)$ credibility matrix, $Z_{1,2}$, for the combined experiment,

$$Z_{1,2} = \Delta\begin{bmatrix}X_1' & X_2'\end{bmatrix}\begin{bmatrix}E_{11} + X_1\Delta X_1' & X_1\Delta X_2'\\ X_2\Delta X_1' & E_{22} + X_2\Delta X_2'\end{bmatrix} , \tag{7.3}$$

which is then used in the estimate:

$$\mathscr{E}\{\beta(\tilde{\theta})\,|\,y_1,X_1;y_2,X_2\} \approx f_\beta(y_1,X_1;y_2,X_2) = \left(I_k - Z_{1,2}\begin{bmatrix}X_1\\ X_2\end{bmatrix}\right)b + Z_{1,2}\begin{bmatrix}y_1\\ y_2\end{bmatrix} . \tag{7.4}$$

If we define individual Z_β-type matrices for each of the experiments individually,

$$Z_i = \Delta X_i' (E_{11} + X_i \Delta X_i')^{-1} \quad , \qquad (i = 1,2) \, , \tag{7.5}$$

then the combined credibility matrix can be written in a simpler form:

$$Z_{1,2} = \begin{bmatrix} Z_1 & Z_2 \end{bmatrix} \begin{bmatrix} I_{n_1} & X_1 Z_2 \\ X_2 Z_1 & I_{n_2} \end{bmatrix}^{-1} \tag{7.6}$$

$$= \begin{bmatrix} Z_1 & Z_2 \end{bmatrix} \begin{bmatrix} (I_{n_1} - X_1 Z_2 X_2 Z_1)^{-1} & -(I_{n_1} - X_1 Z_2 X_2 Z_1)^{-1} X_1 Z_2 \\ -(I_{n_2} - X_2 Z_1 X_1 Z_2)^{-1} X_2 X_1 & (I_{n_2} - X_2 Z_1 X_1 Z_2)^{-1} \end{bmatrix} .$$

Further simplification requires the assumption of full rank for X_1 and X_2, and the definitions (see (4.14) (4.15):

$$\varepsilon_i^{-1} = X_i' E_{ii}^{-1} X_i \quad ; \qquad z_i = \Delta (\Delta + \varepsilon_i)^{-1} \quad ; \qquad (i = 1,2). \tag{7.7}$$

After repeated use of (4.12) and (4.16), the result finally simplifies to

$$Z_{1,2} = \left[(I_k - z_2)(I_k - z_1 z_2)^{-1} z_1 \;\; (I_k - z_1)(I_k - z_2 z_1)^{-1} z_2 \right]. \tag{7.8}$$

Defining the individual classical estimators for each experiment

$$\hat{\beta}_i (y_i) = \varepsilon_i X_i' E_{ii}^{-1} y_i \quad , \qquad (i = 1,2) \quad , \tag{7.9}$$

we obtain finally the combined-experiment estimate,

$$f_\beta (y_1, X_1; y_2, X_2) = (I_k - z^{(1)} - z^{(2)}) b + z^{(1)} \hat{\beta}_1 (y_1) + z^{(2)} \hat{\beta}_2 (y_2) \quad , \tag{7.10}$$

where

$$z^{(1)} = (I_k - z_2)(I_k - z_1 z_2)^{-1} z_1; \quad z^{(2)} = (I_k - z_1)(I_k - z_2 z_1)^{-1} z_2. \tag{7.11}$$

This formula can then be rearranged so as to display a new prior mean, $b^{(2)}$, which is used as input to the second experiment, together with the credibility matrix $z^{(2)}$, in the "single-stage" formula

$$f_\beta(y_1, X_1; y_2, X_2) = (I_k - z^{(2)})b^{(2)} + z^{(2)}\hat{\beta}_2(y_2) \quad . \tag{7.12}$$

Then, we find that

$$\begin{aligned} b^{(2)} &= b + (I_k - z^{(2)})^{-1} z^{(1)} [\hat{\beta}_1(y_1) - b] \\ &= (I_k - z_1)b + z_1\hat{\beta}_1(y_1) = f_\beta(y_1, X_1) \quad , \end{aligned} \tag{7.13}$$

is just the usual first-stage credibility prediction (4.2) or (4.17), which becomes the mean input for the second experiment.

We may further clarify (7.12) by seeing what equivalent regression coefficient covariance, say $\Delta^{(2)}$, is used as input to the second experiment to find the credibility coefficient in the usual way as

$$z^{(2)} = \Delta^{(2)}(\Delta^{(2)} + \varepsilon_2)^{-1} \quad . \tag{7.14}$$

We find

$$\Delta^{(2)} = (\varepsilon_1^{-1} + \Delta^{-1})^{-1} = z_1\varepsilon_1 = \Phi_\beta(X_1) \quad , \tag{7.15}$$

which is just the preposterior estimate of the error covariance (5.2) after the first experiment!

To summarize, we can view the two experiments (n_1, X_1, y_1) (n_2, X_2, y_2):

(1) Either as a combined experiment in which the prior information b and Δ is used in (7.10) to form an estimate of $\beta(\theta)$;

(2) Or as a two-stage process in which b and Δ are used in the first experiment to form $f_\beta(y_1,X_1)$ and $\Phi_\beta(X_1)$, and these values are then used as the prior vector mean and matrix covariance of the regression coefficients for the independent second experiment, forming an estimate of $\beta(\theta)$ using (7.12)(7.14).

The extension to multiple cascaded experiments is obvious. Also, it follows that, prior to both experiments, our estimate of the final covariance matrix is

$$\Phi_\beta(X_1,X_2) = (\varepsilon_2^{-1} + \Phi_\beta^{-1}(X_1))^{-1} = (\Delta^{-1} + \varepsilon_1^{-1} + \varepsilon_2^{-1})^{-1} \quad .$$

In other words, the total final precision is estimated, prior to any experiment, to be the sum of the prior precision plus the observation precision of each experiment.

We now examine several special cases of interest.

7.1 Imprecise Experimental Results

If the first experiment is performed under poor observational conditions, we expect the diagonal elements of E_{11} to be large compared to those of $X_1 \Delta X_1'$. Under these conditions, $z_1 \to 0$, $z^{(2)} \to z_2$, and the results of the first experiment are ignored, with b and Δ used directly as inputs to the second stage. Similar remarks apply to imprecise results in the second experiment; and, of course, if both experiments have high observational variances, then the best forecast is just b.

7.2 Diffuse Prior Information

If, on the other hand, the prior variances of the regression coefficients are very large compared to the imputed covariances ε_1 and ε_2 due to observational error, then z_1 and z_2 approach unity, and we see from (7.14) (7.15), or by careful limits in (7.11), that $z^{(i)} \to (\varepsilon_1^{-1} + \varepsilon_2^{-1})^{-1}\varepsilon_i^{-1}$, (i = 1,2), and

$$f_\beta(y_1,X_1;y_2,X_2) = (\varepsilon_1^{-1} + \varepsilon_2^{-1})^{-1}\left[\varepsilon_1^{-1}\hat{\beta}_1(y_1) + \varepsilon_2^{-1}\hat{\beta}_2(y_2)\right] \quad . \quad (7.16)$$

In other words, the prior information is ignored as the diagonal elements of Δ become large (the prior becomes "diffuse"), and the resulting estimate weights the classical estimators from each experiment in the familiar proportional-to-precision manner. A formula similar to (7.16) is given by sampling theory arguments in the "mixed-estimation" method of Goldberger and Theil [8, Section 5-6][9][27][28].

Alternatively, we may regard this case as one in which a prior mean $\hat{\beta}_1(y_1)$ and a prior covariance ε_1 are used as input to the second experiment.

7.3 Direct Estimate of Regression Parameters

If the first experiment provides a direct measurement of the regression parameters, $\beta(\theta)$, then $n_1 = k$, $X_1 = I_k$, and for consistency, we could call $y_1 = b_1$ a new estimate of b, with covariance of observation errors, $\varepsilon_1 = \Delta_1$, say. Then, the credibility matrix in this special first experiment is $z_1 = \Delta(\Delta + \Delta_1)^{-1}$, the mean input (7.13) to the second experiment is

$$b^{(2)} = (\Delta^{-1} + \Delta_1^{-1})^{-1}\left[\Delta^{-1}b + \Delta_1^{-1}b_1\right] \quad , \tag{7.17}$$

and the covariance matrix input (7.15) is

$$\Delta^{(2)} = (\Delta^{-1} + \Delta_1^{-1})^{-1} \quad . \tag{7.18}$$

In other words, if there are two prior estimates of the regression parameters, then they should be combined in the usual proportional-to-precision manner, and then used as input.

7.4 Similar Experiments

If the design matrix, X, of the two experiments is the same, then the common $z = \Delta(\Delta + \varepsilon)^{-1}$, with $\varepsilon^{-1} = X'E^{-1}X$, and the forecast (7.10) can be written

$$f_\beta(y_1;y_2;X) = \varepsilon(2\Delta + \varepsilon)^{-1}b + 2\Delta(2\Delta + \varepsilon)^{-1}\left[\tfrac{1}{2}(\hat{\beta}(y_1) + \hat{\beta}(y_2))\right] \quad , \tag{7.19}$$

with an obvious definition of the common function $\hat{\beta}(y)$. In this form, the analogy with the many-sample credibility forecast (3.11)(3.12) is obvious, and the extension to t *similar*

experiments is immediate:

$$f_\beta(y_1;y_2;\ldots y_t;X) = [I_k - z(t)]b + z(t)\left[\frac{1}{t}\sum_{i=1}^{t}\hat{\beta}(y_i)\right] \quad , \qquad (7.20)$$

with a new credibility matrix

$$z(t) = t\Delta(t\Delta + \varepsilon)^{-1} \quad . \qquad (7.21)$$

7.5 Repeated Dissimilar Experiments

For completeness, we give the general formulae corresponding to (7.10)(7.11), when t *dissimilar* experiments $(n_1,X_1,y_1)(n_2,X_2,y_2)\ldots(n_t,X_t,y_t)$ are performed. In an obvious extension of notation,

$$f_\beta((y_i,X_i);i=1,2,\ldots,t) = \left[I_k - \sum_{i=1}^{t} z^{(i)}\right]b + \sum_{i=1}^{t} z^{(i)}\hat{\beta}_i(y_i) \quad , \qquad (7.22)$$

where the $z^{(i)}$ are the solutions of

$$\left[z^{(1)}z^{(2)}\ldots z^{(t)}\right] = [1\ 1\cdots 1]\begin{bmatrix} z_1^{-1} & I_k & \cdots & I_k \\ I_k & z_2^{-1} & \cdots & I_k \\ \vdots & \vdots & & \vdots \\ I_k & I_k & & z_t^{-1} \end{bmatrix}^{-1} \quad . \qquad (7.23)$$

The prior-to-experiments estimate of the final covariance of the estimator error is

$$\Phi_\beta(X_1,X_2,\ldots,X_t) = \left(\Delta^{-1} + \sum_{i=1}^{t}\varepsilon_i^{-1}\right)^{-1} \quad ; \qquad (7.24)$$

that is, the final precision is estimated to be the sum of the prior precision plus all of the observational precisions. Of course, as indicated earlier, it is probably easier to compute (7.22) in the recursive manner suggested earlier in this section.

8. Related Work

There are two papers which originated the application of credibility theory to regression problems. In a multidimensional model, with elaborate notation based on practical considerations, Hachemeister [10] has given prediction formulae equivalent to (4.18)(4.19); however, his derivation appears to require the assumption of heteroscedastic error terms, i.e.

$$\Sigma(\theta) = \sigma^2(\theta) I_n \quad , \tag{8.1}$$

or of the sample-mean generalization in which the ith diagonal term of $\Sigma(\theta)$ is $\sigma^2(\theta)/P_i$, where P_i is the "volume" of the ith sample.

He also gives a credibility result for a homogeneous estimator, i.e., with $z_{i0} = 0$ in (3.4), and the remaining credibility coefficients constrained to give an unbiased estimator. For models of this type, one usually has collateral data [17] from similar experiments performed on other risks, with independent values of θ.

Taylor's first paper [25] concentrates on the two-parameter, homogeneous estimator model, using essentially the same assumptions as Hachemeister, but with a simplified unbiasedness constraint. In a later paper [26], Taylor generalizes both the homogeneous and inhomogeneous versions of (4.18) to Hilbert spaces, and shows various special cases.

Turning to exact Bayesian regression results based upon multinormal likelihoods, Raiffa and Schlaiffer [22] give formulae equivalent to (4.17) for the cases in which (1) $\sigma^2(\theta) = \sigma^2$ is a known constant, and the prior on $\beta(\theta)$ is multinormal (b,Δ); (2) $(\sigma^2(\theta),\beta(\theta))$ are inverse-Gamma-multinormally distributed. Other models by Tiao, Zellner, and Chetty [29][30][32][34] concentrate on the use of a diffuse prior density, $p(\beta,\sigma^2) \propto \sigma^{-1}$, or its multidimensional equivalent [32, Chapter 8]; thus, after one experiment, $\hat{\beta}_1(y_1)$ is "fully credible," or after two experiments, results similar to (7.16) are obtained. Of course, since these are exact Bayesian results, the complete posterior distributions of the parameter are available--usually some variation of the multivariate-t density.

In [32, p. 240], Zellner takes an "informative" prior which is slightly more general than the usual natural-conjugate prior for the multinormal; his likelihood is multivariate, with homeoscedastic errors, which can be reinterpreted as

single-variate with arbitrary $\Sigma(\theta)$. By expanding the resulting posterior density for the regression parameters, he finds from the leading normal term a mean estimate which is "a 'matrix weighted average' of the prior mean...and the least-squares quantity $\hat{\beta}$ whose weights are the inverse of the prior covariance C and the sample covariance matrix." This is, of course, just our result (4.17)(4.18)(5.2), gotten as an approximation for arbitrary likelihood and prior densities.

We have also indicated that, using sampling theory arguments, Goldberger and Theil [8][9][27][28] have obtained formulae similar to (7.16), except that, since $\sigma_i^2(\theta)$ $(i = 1,2)$ in ε_1, ε_2 are unknown, they propose substituting various reasonable sample estimates.

9. Exact Results

It can be seen from the above that the credibility formulae presented here are exact when the likelihood is multinormal, and the prior is from a natural conjugate family. However, there are additional cases in which the credibility results are exact, based upon the Koopmans-Pitman-Darmois exponential-type families, and their (suitably enriched) natural conjugate priors. (See [12][13][14] for exact results for the model of (3.11).) These will be reported in a separate paper.

10. Extensions

Many of the topics which are considered as extensions in classical works on regression are already covered by our basic model, since no special assumption about the error covariance matrix $\Sigma(\theta)$ has been made; for example, error terms may be autocorrelated. Multivariate regression models are already "serially" included, and it remains only to translate them into the usual "parallel" notation. And, by following the discussion in Section 6, a variety of random input models may be elaborated; for example, successive inputs may follow a "random shocks" process [15].

There are many interesting regression models in which the design matrix is not of full rank. In these cases, (4.2) and (4.4) are still viable, even though the classical estimators do not exist. Or one may add additional constraints, based upon external considerations, until the problem is "identifiable," in the classical sense. The particular problem of estimating flows in a network will be the topic of a future report.

For a simple linear regression, one can also talk about problems of inverse regression; that is, given y, what was

the input x? These questions arise in various problems of measurement, and a detailed study of instrument calibration and measurement using credibility methods may be found in [18].

BIBLIOGRAPHY

[1] Ando, A. and Kaufman, G.M. "Bayesian Analysis of the Independent Multinormal Process--Neither Mean Nor Precision Known." J. Amer. Statist. Assoc., 60, pp. 347-358 (1965).

[2] Bodewig, E. Matrix Calculus (2nd Edition). North-Holland, Amsterdam (1959).

[3] Box, G.E.P. and Tiao, G.C. Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading, Massachusetts (1973).

[4] Bühlmann, H. "Experience Rating and Credibility." ASTIN Bulletin, 4, Part 3, pp. 199-207 (July, 1967).

[5] _____ Mathematical Methods in Risk Theory. Springer-Verlag, New York (1970).

[6] Cox, D.R. and Hinkley, D.V. Theoretical Statistics. Chapman and Hall, London (1974).

[7] Florens, J.-P., Mouchart, M. and Richard, J.-F. "Bayesian Inference in Error-in-Variables Models." J. of Multivariate Analysis, 4, No. 4, pp. 419-452. (1974).

[8] Goldberger, A.S. Econometric Theory. J. Wiley & Sons, New York (1964).

[9] _____ "Efficient Estimation in Overidentified Models: An Interpretive Analysis." Chapter 7 in Structural Equation Models in the Social Sciences, A.S. Goldberger and O.D. Duncan (Eds.), Seminar Press, New York (1973).

[10] Hachemeister, C.A. "Credibility for Regression Models with Application to Trend." Proceedings of Actuarial Research Conference on Credibility Theory, Berkeley, California, September, 1974. Academic Press, New York (1975).

[11] Jewell, W.S. "The Credible Distribution." ORC 73-7, Operations Research Center, University of California, Berkeley (August, 1973). ASTIN Bulletin, 7, Part 3, pp. 237-269 (March, 1974).

[12] _____ "Credible Means are Exact Bayesian for Simple Exponential Families." ORC 73-21, Operations Research Center, University of California, Berkeley (October, 1973). ASTIN Bulletin, 8, Part 1, pp. 77-90 (September, 1974).

[13] _____ "Exact Multidimensional Credibility." ORC 74-14, Operations Research Center, University of California, Berkeley (May, 1974). Mitteilungen der Vereinigung Schweizerischer Versicherungs-mathematiker, 74, No. 2, pp. 193-214 (1974).

[14] _____ "Regularity Conditions for Exact Credibility." ORC 74-22, Operations Research Center, University of California, Berkeley (July, 1974). To appear in ASTIN Bulletin.

[15] _____ "Model Variations in Credibility Theory." ORC 74-25, Operations Research Center, University of California, Berkeley (August, 1974). Proceedings of Actuarial Research Conference on Credibility Theory, Berkeley, California, September, 1974. Academic Press, New York (1975).

[16] _____ "Two Classes of Covariance Matrices Giving Simple Linear Forecasts." RM-75-17, International Institute for Applied Systems Analysis, Laxenburg, Austria (May, 1975). To appear in Scandinavian Actuarial Journal.

[17] _____ "The Use of Collateral Data in Credibility Theory: A Hierarchical Model." RM-75-24, International Institute for Applied Systems Analysis, Laxenburg, Austria (June, 1975). To appear in Giornale dell' Istituto Italiano degli Attuari.

[18] Jewell, W.S. and Avenhaus, R. "Bayesian Inverse Regression and Discrimination: An Application of Credibility Theory." RM-75-27, International Institute for Applied Systems Analysis, Laxenburg, Austria (June, 1975).

[19] Lindley, D.V. and Smith, A.F.M. "Bayes Estimates for the Linear Model." J. Royal Statist. Soc., (B), 34, pp. 1-41 (1972).

[20] Malinvaud, E. Statistical Methods of Econometrics (2nd Revised Edition). North-Holland, Amsterdam (1970).

[21] Morales, J.A. Bayesian Full Information Structural Analysis. Springer-Verlag, Berlin (1971).

[22] Raiffa, H. and Schlaiffer, R. Applied Statistical Decision Theory. Harvard Business School, Boston (1961).

[23] Rao, C.R. Linear Statistical Inference and its Applications. J. Wiley & Sons, New York (1965).

[24] Rothenberg, T.J. Efficient Estimation with A Prior Information. Yale University Press, New Haven, Connecticut (1973).

[25] Taylor, G.C. "Credibility for Time-Heterogeneous Loss Ratios." Research Paper No. 55, MacQuarie University, Sydney, July, 1974. Proceedings of Actuarial Research Conference on Credibility Theory, Berkeley, California, September, 1974. Academic Press, New York (1975).

[26] _____ "Abstract Credibility." MacQuarie University, Sydney, and Herriot-Watt University, Edinburgh (February, 1975).

[27] Theil, H. "On the Use of Incomplete Prior Information in Regression Analysis." J. Amer. Statist. Assoc., 58, pp. 401-414 (1963).

[28] Theil, H. and Goldberger, A.S. "On Pure and Mixed Statistical Estimation in Economies." Intern. Econ. Rev., 2, pp. 65-78 (1961).

[29] Tiao, G.C. and Zellner, A. "Bayes Theorem and the Use of Prior Knowledge in Regression Analysis." Biometrika, 51, pp. 219-230 (1964).

[30] Tiao, G.C. and Zellner, A. "On the Bayesian Estimation of Multivariate Regression." J. Royal Statist. Soc., (B), 26, pp. 277-285 (1964).

[31] Tocher, K.D. "Discussion on Mr. Box and Dr. Wilson's Paper." J. Royal Statist. Soc., (B), 13, pp. 39-42 (1951).

[32] Zellner, A. An Introduction to Bayesian Inference in Econometrics. J. Wiley & Sons, New York (1971).

[33] _____ "Bayesian Analysis of Regression Error Terms." J. Amer. Statist. Assoc., 70, pp. 138-144 (1975).

[34] Zellner, A. and Chetty, V.K. "Prediction and Decision Problems in Regression Models from the Bayesian Point of View." J. Amer. Statist. Assoc., 60, pp. 608-616 (1965).

RM - 75 - 27

INTERNATIONAL INSTITUTE FOR IIASA APPLIED SYSTEMS ANALYSIS

RESEARCH MEMORANDUM

BAYESIAN INVERSE REGRESSION AND DISCRIMINATION: AN APPLICATION OF CREDIBILITY THEORY

R. Avenhaus

W. S. Jewell

June 1975

SCHLOSS LAXENBURG
2361 Laxenburg
AUSTRIA

RM-75-27

BAYESIAN INVERSE REGRESSION AND DISCRIMINATION: AN APPLICATION OF CREDIBILITY THEORY

R. Avenhaus

W.S. Jewell

June 1975

Bayesian Inverse Regression and Discrimination: An Application of Credibility Theory

R. Avenhaus and W.S. Jewell

Abstract

Many measurement problems can be formulated as follows: a certain linear relationship between two variables is to be estimated by using pairs of input and output data; the value of an unknown input variable is then estimated, given an observation of the corresponding output variable. This problem is often referred to as inverse regression or discrimination.

In this paper, we formulate a general Bayesian calibration and measurement model for this problem, in which prior information is assumed to be available on the relationship parameters, the possible values of the unknown input, and the output observation error. Simplified and easily interpreted formulae for estimating the posterior mean and variance of the input are then developed using the methods of credibility theory, a linearized Bayesian analysis developed originally for insurance estimation problems. A numerical example of the calibration of a calorimeter to measure nuclear material is given.

1. Problem Formulation

In this paper, we consider problems of the following kind: we wish to estimate the value of a certain state variable x which cannot be measured directly, or only with very large error or effort. We know, however, of another state variable y, which is statistically dependent on x, and which can be measured more easily or accurately. Thus, in principle, we can estimate the relationship between x and y, and then, with small effort, obtain x by measuring y and using the inverse relationship.

However, difficulty arises because we must use other pairs,

(x_i, y_i) $(i = 1,2,\ldots,n)$, to estimate the relationship. Often these will have been determined for other objectives and under different experimental conditions. Thus, the true values of independent and dependent variables may not be precisely known, or the relationship itself may be slightly different than it appears from the data.

Finally, as in most physical problems, we assume that a great deal of collateral information is available which gives us some prior idea of relationship between x and y, and even of the unknown value x we are trying to estimate. In other words, we wish to make a Bayesian formulation of the problem.

Three examples of this class of problem are given below.

A. Calibration and Indirect Measurement of Nuclear Materials

Nuclear materials, e.g. plutonium, are extremely difficult to measure directly by chemical means. Therefore, one uses indirect methods, based upon the heat production or the number of neutrons emitted, in order to estimate the amount of material present. From well-known physical laws, we have a general relationship between these variables, but any measurement instrument based on these principles needs first to be calibrated. Usually, this calibration can be done with the aid of standard inputs, containing known amounts of nuclear materials. However, these inputs (x_i) are not generally under our control, and in some cases, may have residual

imprecisions in their values.

Measurement instruments often have longer-term drifts, during which they tend to loose their original calibration. For this reason, measurement of a given production run often consists of two distinct phases: (re)calibration of the instrument, and actual indirect measurement. With a fixed amount of time available, it is of interest to determine how much time should be spent on the two phases, assuming that additional time spent on each observation reduces observational error.

B. Estimation of Family Incomes by Polling

We wish to estimate, through a public opinion poll, the distribution of family incomes in a certain city district. As the major part of the population will not be willing to divulge their incomes, or will give only a very imprecise figure, we look for a dependent variable which can be more easily determined. According to the literature (see, e.g. [10]), housing expenses are strongly related to family income, and, furthermore, it may be assumed that the population is less reluctant to divulge this figure, even though they may not be able to do so precisely. Clearly, to determine this relationship exactly, we must have some families in this district who are willing to give both their total income and their household expenses. On the other hand, we have strong prior information on this relationship from similar surveys, and may have general information

on income distribution from census and other sources.

C. Missing Variables in Bayesian Regression

In a paper with this title [11], Press and Scott consider a simple linear regression problem in which certain of the independent variables, x_i, are assumed to be missing in a nonsystematic way from the data pairs (x_i, y_i). Then, under special assumptions about the error and prior distributions, they show that an optimal procedure for estimating the linear parameters is to first estimate the missing x_i from an inverse regression based only on the complete data pairs.

Problems of this kind are described in textbooks on the theory of measurements, and are sometimes called *discrimination problems* (Brownlee [1], Miller [9]).

In the following, we shall formulate these problems as *Bayesian calibration and measurement* problems, in the sense of Dunsmore [3] [4], Hoadley [5], and Lindley [8]. This formulation is quite general, and although the language corresponds to that of example A, the translation to other examples is easily made.

Because of the strong distributional specification requirements of the full Bayesian analysis, we shall then use the approach of *credibility theory* to find best linear approximations to moments of interest. The resulting formulae enable us to easily display the relative value of prior information, on the one hand, and information obtained in the calibration, on the other. We will develop further the optimization problem

described in Example A above, and will consider a numerical example of calibration and indirect measurement of nuclear material.

2. Bayesian Calibration and Measurement Model

To develop the Bayesian model, we suppose that:

(1) *Calibration* consists of n independent pairs of input and output observations $(\underline{x},\underline{y}) = ((x_i,y_i),\ i = 1,2,\ldots,n)$. ($x_i$ is a relatively precise or *standard input*, and y_i is the *observed output* on a *measurement instrument*, which specifies a statistical relationship between these pairs through a conditional *measurement density*, $p(y_i|x_i,\theta)$; the measurement density depends upon a fixed but unknown *measurement parameter* θ, for which we have a prior density, $p(\theta)$);*

(2) *Measurement* consists of using the same instrument on a sample of *unknown input*, $\tilde{x} = x$, to obtain an output $\tilde{y} = y$, say; the problem is then to *infer* the value of x. Since this cannot be accomplished, we must, in general, settle for an estimate, $\hat{x}$, which, in the remainder of the paper, we will assume to be $\mathscr{E}\{\tilde{x}|y;\underline{x},\underline{y}\}$. Other Bayes estimators may be important in other physical situations.

Following [8], we see that we must compute the posterior conditional density,

---

*We use the convention that the arguments of any $p(\cdot)$ indicate the particular density in question, which may be with respect to Lebesgue or discrete measure. Where necessary, we indicate a random variable with a tilde; i.e., $\tilde{x}$ is the random variable corresponding to x, etc..

$$p(x|y;\underline{x},\underline{y}) = \frac{p(x,y;\underline{y}|\underline{x})}{p(y;\underline{y}|\underline{x})}$$

$$= \frac{\int p(y,\underline{y}|x,\underline{x},\theta)\, p(\theta|x,\underline{x})\, p(x|\underline{x},\theta)\, d\theta}{\int p(x',y,\underline{y}|\underline{x})\, dx'} \qquad (2.1)$$

from which the mean, $\mathcal{E}\{\tilde{x}|y;\underline{x},\underline{y}\}$, will be our estimate of the unknown input, and the variance, $\mathcal{V}\{\tilde{x}|y;\underline{x},\underline{y}\}$, will be the norm for our optimization problem, since we wish to make the estimate as precise as possible in the least-squares sense.

To proceed further, we must make additional statistical assumptions appropriate to our problem:

(1) Given θ, we assume that the measurements are independent:

$$p(y,\underline{y}|x,\underline{X},\theta) = p(y|x,\theta) \prod_{i=1}^{n} p(y_i|x_i,\theta) \quad ;$$

(2) We assume that the prior on the measurement parameter is unrelated to any of the inputs:

$$p(\theta|x,\underline{x}) = p(\theta) \quad ;$$

(3) Any unknown input in the measurement process, x, is selected independently from the standard inputs, $\underline{x} = [x_1, x_2, \ldots, x_n]'$ and the parameter θ:

$$p(x|\underline{x},\theta) = p(x) \quad .$$

The third assumption is the strongest, and may not hold, for example, when the calibration inputs and the test input come from the same production process. However, in our case, we assume that the calibration inputs are independent standards.

By elementary manipulations, we obtain:

$$p(x|y;\underline{x}.\underline{y}) = \frac{p(x) \int p(y|x,\theta)\, p\,(\theta|\underline{x},\underline{y})\, d\theta}{\int p(y|\theta')\, p\,(\theta'|\underline{x},\underline{y})\, d\,\theta'} \quad , \tag{2.2}$$

where

$$p(\theta|\underline{x},\underline{y}) = \frac{\prod_{i=1}^{n} p\,(y_i|x_i,\theta)\, p\,(\theta)}{\int \prod_{j=1}^{n} p\,(y_j|x_j,\theta')\, p\,(\theta')\, d\theta'} \quad . \tag{2.3}$$

Notice that the denominators of (2.2) and (2.3) are just normalizations, which may be computed directly at any time.

In the above form, it is clear that the problem breaks apart mathematically into two problems:

(1) The updating of $p(\theta)$ to $p(\theta|\underline{x},\underline{y})$ (calibration);

(2) The calculation of moments of interest for $p(x|y,\theta)$, averaged over the appropriate density of θ measurement.

We tackle these problems in reverse order, since the only effect of calibration is to modify the prior information about the regression parameters and to improve the precision of this estimate.

3. Estimation of Input Using Credibility Theory

To find the moments of $p(x|y,\theta) = p(y|x,\theta)\, p\,(x)/\int p(y|x',\theta)\, p(x')\, dx'$, we must in the general case make distributional assumptions about $p(x)$ and $p(y|x,\theta)$. However, since only the moments of this density are of interest, it is desirable to

have a simpler, distribution-free approach, such as that provided by *credibility theory* [6][7]. In this approach, Bayesian means conditional on given data w, say, are approximated by linear combinations of certain functions of w, chosen from physical considerations; the coefficients are then chosen to minimize the mean-square approximation error prior to w. In certain cases, these approximation formulae are also the exact Bayesian conditional means [6].

The usual assumption about a measurement process is that, given the measurement parameter θ, there is a linear relation between the true input and the true output, but that the observed process may contain an additional uncorrelated measurement observation error, with zero mean and known variance. This may be conveniently expressed as:

$$\mathscr{E}\{\tilde{y}|x,\theta\} = \beta_1(\theta) + \beta_2(\theta)\, x \quad ; \tag{3.1}$$

$$\mathscr{V}\{\tilde{y}|x,\theta\} = \sigma_M^2 \quad . \tag{3.2}$$

(In other applications, the observation error may also depend upon θ or the level of x.) We call $\beta_1(\theta)$, $\beta_2(\theta)$ the *instrument parameters*.

We know that, for general $p(x,y|\theta)$, the fact that the regression of y upon x (3.1) is linear does not necessarily mean that the regression of x upon y is linear in y. However, it is true in the case of the normal and some other bivariate distributions, and seems a desirable characteristic of any measurement process. Therefore, we shall assume that our prior estimate of the true input x, given an observed output y, may

be approximated by the linear function:

$$\mathscr{E}\{\tilde{x}|y\} = \mathscr{E}\mathscr{E}\{\tilde{x}|y,\tilde{\theta}\} \approx f(y) = z_0 + z_1 y \quad , \tag{3.3}$$

where the "credibility coefficients" z_0, z_1 are chosen so as to minimize the approximation error variance:

$$H_A = \mathscr{E}\left[\mathscr{E}\{\tilde{x}|\tilde{y}\} - f(\tilde{y})\right]^2 \quad . \tag{3.4}$$

For the remainder of this section, we shall treat the averaging over θ as if it were with respect to the prior $p(\theta)$, realizing that in the next section we shall change to $p(\theta|\underline{x},\underline{y})$, to add the information provided by the calibration.

One can easily show [6,7][2, Appendix 3] that the optimal credibility coefficients are given by:

$$z_0 = \mathscr{E}\{\tilde{x}\} - z_1 \mathscr{E}\{\tilde{y}\} \quad ; \tag{3.5}$$

$$z_1 = \frac{\mathscr{C}\{\tilde{y};\tilde{x}\}}{\mathscr{V}\{\tilde{y}\}} \quad ; \tag{3.6}$$

so that the optimal estimator is unbiased.

$\mathscr{E}\{\tilde{x}\}$ represents our prior estimate of the value of the input to be measured; the remaining moments must be calculated from our measurement assumptions (3.1) (3.2). From (3.1):

$$\mathscr{E}\{\tilde{y}\} = b_1 + b_2 \mathscr{E}\{\tilde{x}\} \quad , \tag{3.7}$$

where

$$b_i = \mathscr{E}\{\beta_i(\tilde{\theta})\} \qquad (i = 1,2) \tag{3.8}$$

are the mean prior estimates of the instrument parameters. By unconditioning (3.2) on x and θ, we find:

$$\mathcal{V}\{\tilde{y}\} = \sigma_M^2 + \mathcal{V}\{\tilde{x}\}\,(b_2^2 + \Delta_{22}) + \Delta_{11} + 2\Delta_{12}\mathcal{E}\{\tilde{x}\} + \Delta_{22}\left[\mathcal{E}\{\tilde{x}\}\right]^2 \quad ,$$

where

$$\Delta_{ij} = \mathcal{C}\{\beta_i(\tilde{\theta})\,;\,\beta_j(\tilde{\theta})\} \qquad (i,j = 1,2) \tag{3.10}$$

are the prior estimates of the (co)variances in the instrument parameters. We see that the total prior-to-measurement variance in the observation is composed of three groups of terms:

(1) The observation error variance;

(2) The prior variation in input;

(3) (Co)variances in instrument parameters.

An increase in any one of these will reduce the weight, z_1, attached to the observed output, y, in (3.3).

There is only one prior source of covariance between input and output:

$$\mathcal{C}\{\tilde{y};\tilde{x}\} = b_2\mathcal{V}\{\tilde{x}\} \quad , \tag{3.11}$$

which means that, as the uncertainty in the input increases, one must attach more importance to the observed output in (3.3).

For convenience, we reproduce the final formula for the estimate of the true input:

$$f(y) = \mathcal{E}\{\tilde{x}\} + z_1(y - \mathcal{E}\{\tilde{y}\}) = (1 - b_2 z_1)\,\mathcal{E}\{\tilde{x}\} + z_1(y - b_1) \quad ;$$

$$z_1 = \frac{b_2\mathcal{V}\{\tilde{x}\}}{\sigma_M^2 + \mathcal{V}\{\tilde{x}\}\,(b_2^2 + \Delta_{22}) + \Delta_{11} + 2\Delta_{12}\mathcal{E}\{\tilde{x}\} + \Delta_{22}\left[\mathcal{E}\{\tilde{x}\}\right]^2} \quad . \tag{3.13}$$

Thus, in the credibility approach, only seven prior moments must

be specified: the mean and variance of the potential input, and the two means and three (co)variances of the instrument coefficients.

It is of interest to examine several limiting cases of the estimator (3.12) (3.13) in more detail. First, as already mentioned, if either the observation error variance σ_M^2 or any of the instrument variances is very large (sometimes called a "diffuse" calibration prior), then, since z_1 vanishes, the best estimate of $\tilde{x}$ is its prior mean, $\mathscr{E}\{\tilde{x}\}$; the measurement process gives little additional information. Similarly, the vanishing of $\mathscr{V}\{\tilde{x}\}$ makes $\mathscr{E}\{\tilde{x}\}$ very reliable.

On the other hand, suppose that we have a "diffuse" prior on the level of input, that is, although $\mathscr{E}\{\tilde{x}\}$ is given, $\mathscr{V}\{\tilde{x}\} \to \infty$. In this case the forecast can be rewritten:

$$f(y) = \left[1 + (\Delta_{22}/b_2^2)\right]^{-1} \left[(\Delta_{22}/b_2^2)\mathscr{E}\{\tilde{x}\} - (b_1/b_2) - (y/b_2)\right] \quad . \quad (3.14)$$

If Δ_{22}/b_2^2 is small compared with unity, we obtain exactly the deterministic result corresponding to (3.11), $y = b_1 + b_2 x$.

In the optimization model of Section 6, we shall need the mean-square value of the error between the true value x and the predictor f(y), that is, the *variance of forecast error*:

$$H = \mathscr{E}\{(\tilde{x} - f(\tilde{y})^2\} \quad . \quad (3.15)$$

But, by elementary manipulations,

$$H = H_O + H_A \quad , \quad (3.16)$$

where H_O is the *irreducible forecast variance using the Bayesian conditional mean*:

$$H_O = \mathscr{E}\mathscr{E}\{(\tilde{x} - \mathscr{E}\{\tilde{x}|\tilde{y}\})^2|\tilde{y}\} = \mathscr{E}\mathscr{V}\{\tilde{x}|\tilde{y}\} \quad , \tag{3.17}$$

and H_A is given by (3.4).

With the optimal choice of credibility coefficients, we obtain:

$$H = \mathscr{V}\{\tilde{x}\} - z_1\mathscr{C}\{\tilde{y};\tilde{x}\} = \mathscr{V}\{\tilde{x}\}(1 - z_1 b_2) \quad . \tag{3.18}$$

H in (3.15) and (3.18) is the variance of forecast error for one inverse measurement. If r such measurements are performed, with independent, identically distributed inputs, then one can easily show that the variance of the total error will be:

$$H^{(r)} = r\mathscr{V}\{\tilde{x}\}(1 - z_1 b_2) + (r^2 - r) z_1^2 (\Delta_{11} + 2\Delta_{12}\mathscr{E}\{\tilde{x}\} + \Delta_{22}[\mathscr{E}\{\tilde{x}\}]^2) \quad . \tag{3.19}$$

We see that, in addition to the expected first term which is r times (3.18), there is a component which is proportional to r^2. This represents a possible persistence of error due to instrument parameter covariances, which may cause the individual forecast errors to be positively correlated.

4. Updating of Instrument Parameters Using Credibility Theory

We turn now to the problem of incorporating the results of the calibration experiments into our prior-to-measurement density on θ. Remember that the number, n, of such experiments, and the previously calibrated levels of the inputs, $x_i (i=1,2,\ldots,n)$, are assumed to be fixed by external considerations. See also Section 6 below.

Assuming that (3.1) and (3.2) apply also to calibration (i.e. the same instrument is used), we may write:

$$\mathscr{E}\{\tilde{\underline{y}}|\underline{x},\theta\} = X\,\underline{\beta}(\theta) \quad , \tag{4.1}$$

$$\mathscr{C}\{\tilde{\underline{y}};\tilde{\underline{y}}|\underline{x},\theta\} = \sigma_C^2 I_n \qquad (*) \quad , \tag{4.2}$$

where

$$\tilde{\underline{y}} = [\tilde{y}_1,\tilde{y}_2,\ldots,\tilde{y}_n]' \quad , \quad \underline{x} = [x_1,x_2,\ldots,x_n]' \quad ,$$

$$\underline{\beta}(\theta) = [\beta_1(\theta),\beta_2(\theta)]' \quad , \quad X = [\underline{1}_n,\underline{x}] \quad ,$$

$\underline{1}_n$ is a vector of n ones, I_n is the unit matrix of order n, and σ_C^2 is the observation variance for <u>each</u> output $y_i (i=1,2,\ldots,n)$. We thus have a formulation as a Bayesian regression problem, in which we want to estimate various moments of $p(\underline{\beta}(\theta)|\underline{x},\underline{y})$. In particular, from (3.8)(3.10)(3.13)(3.18), we see that the first and second moments:

$$\mathscr{E}\{\underline{\beta}(\tilde{\theta})|\underline{x},\underline{y}\} \quad ; \quad \mathscr{C}\{\underline{\beta}(\tilde{\theta});\underline{\beta}(\tilde{\theta})|\underline{x},\underline{y}\}$$

will be needed.

(*) Vector covariance is defined as

$$\mathscr{C}\{\tilde{\underline{w}};\tilde{\underline{z}}\} = \mathscr{E}\{\tilde{\underline{w}}\,\tilde{\underline{z}}'\} - \mathscr{E}\{\tilde{\underline{w}}\}[\mathscr{E}\{\tilde{\underline{z}}\}]'$$

for any two random vectors $\tilde{\underline{w}}$ and $\tilde{\underline{z}}$.

Rather than make distributional assumptions, such as those followed in [13], we shall again make a credibility approximation, this time to $\mathscr{E}\{\underline{\beta}(\tilde{\theta}) \mid \underline{x},\underline{y}\}$. The appropriate theory has been developed in [7], and we shall give only the necessary results here.

First, we approximate the desired mean instrument parameter vector by a linear function of the data vector $\underline{y}$:

$$\mathscr{E}\{\underline{\beta}(\tilde{\theta}) \mid \underline{x},\underline{y}\} \approx \underline{g}(\underline{y}) = \underline{z}_0 + Z\underline{y} \quad , \tag{4.3}$$

where $\underline{g}$, $\underline{z}_0$ are two-vectors, Z is a $2 \cdot n$ matrix, and the credibility coefficients are chosen so as to minimize the mean-square approximation of both components to those of the Bayesian conditional mean vector. After some algebra it is shown in [7] that the optimal credibility forecast can be written as:

$$\underline{g}(\underline{y}) = (I_2 - \underline{z})\underline{b} + \underline{z}\,\hat{\underline{\beta}}(\underline{y}) \quad , \tag{4.4}$$

where $\underline{b} = [b_1,b_2]'$ is the vector of prior-to-calibration means, $\underline{z}$ is a new $2 \cdot 2$ *credibility matrix*:

$$\underline{z} = [\Delta(X'E^{-1}X)][I_2 + \Delta(X'E^{-1}X)]^{-1} \tag{4.5}$$

(the terms in square brackets commute), and $\hat{\underline{\beta}}(\underline{y})$ is the classical *regression estimator* of $\tilde{\underline{\beta}}$:

$$\hat{\underline{\beta}}(y) = (X'E^{-1}X)^{-1}\,X'E^{-1}\underline{y} \quad . \tag{4.6}$$

Δ is the $2 \cdot 2$ matrix of prior-to calibration covariances defined in (3.10), and

$$E = \mathscr{E}\mathscr{C}\{\tilde{\underline{y}};\tilde{\underline{y}}\,|\,\underline{x},\ \tilde{\theta}\} = \sigma_C^2\ I_n \quad . \tag{4.7}$$

Thus, in our model, the "regression errors" are "homoscedastic", and we get the further simplifications:

$$\underline{z} = [\Delta X'\underline{X}]\,[\sigma_C^2\ I_2 + \Delta X'\underline{X}]^{-1} \quad , \tag{4.8}$$

and

$$\hat{\underline{\beta}}(\underline{y}) = (X'X)^{-1}X'\underline{y} \quad , \tag{4.9}$$

where

$$XX' = nM = n \begin{bmatrix} 1 & \sum_{i=1}^{n} x_i/n \\ \sum_{i=1}^{n} x_i/n & \sum_{i=1}^{n} x_i^2/n \end{bmatrix} = n \begin{bmatrix} 1 & m_1 \\ m_1 & m_2 \end{bmatrix}, \tag{4.10}$$

i.e. n times a matrix of deterministic moments m_1, m_2 describing the predetermined calibration inputs. One may easily verify that:

$$M^{-1} = \frac{1}{(m_2 - m_1^2)} \begin{bmatrix} m_2 & -m_1 \\ -m_1 & 1 \end{bmatrix} \quad .$$

The results (4.4) (4.8) (4.9) are intuitively very satisfying, for they show that our estimate of the instrument coefficients prior to calibration should be taken as a linear mixture of our prior hypothesis, $\underline{b}$, and of the well-known classical estimator, $\hat{\underline{\beta}}(y)$. The credibility attached to the latter depends upon the so-called *design matrix*, X, the observation error variance, σ_C^2, and the instrument covariances, Δ. (See Jewell [7]).

Several limiting cases are of interest. First, as our observation error variance gets very large, $\underline{z}$ vanishes, and no credibility is attached to the calibration experiment -- it is better to stick with the prior estimates.

Conversely, if all the prior instrument covariances, Δ_{ij}, get very large, then $\underline{z} \to I_2$, and "full credibility" is attached to the calibration data; the same result occurs as $\sigma_C^2 \to 0$. Note also that full credibility occurs as the length of the calibration run, n, increases, as long as the successive inputs are chosen in such a way as to keep m_1 and m_2 about the same; in other words, the more calibration, the more weight is attached to the results.

The above model may be easily generalized to the case where the standard inputs themselves are subject to errors. In this case, we suppose that the selection of a "target input" i specifies $\mathcal{E}\{\tilde{x}_i\}$, rather than x_i; the actual input differs from the mean by a known variance, $\mathcal{V}\{\tilde{x}_i\}$. The reader may easily verify that the above formulae again apply, with $X = [\underline{1}_n, \mathcal{E}\{\underline{\tilde{x}}\}]$ and with (4.7) replaced by a new diagonal matrix, with terms:

$$E_{ii} = \sigma_C^2 + (b_2^2 + \Delta_{22})\, \mathcal{V}\{\tilde{x}_i\} \qquad (i=1,2,\ldots,n) \ , \quad (4.11)$$

In the general case, the formulae (4.5) (4.6) must now be used; however, if the precision of the standards is the same, the regression is again homeoscedastic, and (4.8) (4.9) may be used, but with σ_C^2 replaced by (4.11).

As far as the mean-square error in fitting $\underline{\beta}(\theta)$ by (4.4) is concerned, we can also show that the prior covariance matrix,

with optimal choice of credibility coefficients, is:

$$\begin{aligned}\phi(X) &= \mathscr{E}\{(\underset{\sim}{\beta}(\theta) - \underset{\sim}{g}(\underset{\sim}{y}))(\underset{\sim}{\beta}(\theta) - \underset{\sim}{g}(\underset{\sim}{y}))' | X\} \\ &= (I_2 - \underset{\sim}{z})\Delta = \underset{\sim}{z}(X'E^{-1}X)^{-1} \quad .\end{aligned} \tag{4.12}$$

If this fit is good, then ϕ_{ij} will be a good approximation to $\mathscr{E}\{\beta_i(\tilde{\theta}) ; \beta_j(\tilde{\theta})\}$ after the calibration, at least as we perceive it to be before we actually obtain the outputs $\underset{\sim}{y}$. In other words, $\phi(X)$ is our *preposterior estimate* of the covariance between instrument parameters.

It should be remembered that only the diagonal terms of (4.12) were individually optimized in the choice of credibility coefficients; one can easily show that the diagonal elements of $\phi(X)$ are less than those of Δ.

5. Integration of the Calibration and Measurement Stages

We may now complete our arguments about the relationship between Sections 3 and 4, in light of the knowledge available at each stage of the physical problem.

First, with only a prior hypothesis about our instrument available, and no calibration contemplated, our best estimate of $\underset{\sim}{\beta}(\tilde{\theta})$ is $\underset{\sim}{b}$, with covariance Δ. If an inverse measurement were to be performed at this point, (3.12) (3.13) is the formula we would use to estimate the true input, and H in (3.18) is the estimate now of the variance in this estimate.

Now, suppose we contemplate performing a calibration experiment (X,n), with a fixed number of standards and fixed input design, but the results of the calibration are not yet available.

We still have no basis for revising $\mathscr{E}\{\underline{\beta}(\tilde{\theta})\}$, since the formula (4.4) is, prior-to-calibration, unbiased. However, the knowledge that there will be a calibration will reduce our instrument covariance terms from Δ to $\phi(X)$. Therefore, prior to calibration, our estimate of the forecast error variance after measurement changes from (3.18) to:

$$H(X) = \mathscr{V}\{\tilde{x}\} - \frac{b_2^2\left[\mathscr{V}\{\tilde{x}\}\right]^2}{\sigma_M^2 + \mathscr{V}\{\tilde{x}\}(b_2^2+\phi_{22}) + \phi_{11} + 2\phi_{12}\mathscr{E}\{\tilde{x}\} + \phi_{22}\left[\mathscr{E}\{\tilde{x}\}\right]^2} \quad . \tag{5.1}$$

(This is the point at which optimization of the next section will be carried out). Similar modification applies to (3.19).

We now perform the calibration experiment, obtaining $\underline{y}$ and the revised estimates, $\underline{g}(\underline{y})$, of $\mathscr{E}\{\underline{\beta}(\tilde{\theta})|\underline{y},X\}$ from (4.4). These revised estimates of the instrument coefficients are then used in (3.12) and (3.13), which become:

$$f(y|\underline{y},X) = \left[1 - g_2(\underline{y})z_1(\underline{y},X)\right]\mathscr{E}\{\tilde{x}\} + z_1(\underline{y},X)\left[y - g_1(\underline{y})\right] \quad ; \tag{5.2}$$

$$z_1(\underline{y},X) = \frac{g_2(\underline{y})\mathscr{V}\{\tilde{x}\}}{\sigma_M^2 + \mathscr{V}\{\tilde{x}\}\left\{\left[g_2(\underline{y})\right]^2 + \phi_{22}\right\} + \phi_{11} + 2\phi_{12}\mathscr{E}\{\tilde{x}\} + \phi_{22}\left[\mathscr{E}\{\tilde{x}\}\right]^2} \quad . \tag{5.3}$$

This is the final estimator for any unknown input, after the calibration has been performed.

We admit that it should, in principle, be possible to revise our estimate of the covariance of the instrument coefficients, ϕ, after the actual calibration outputs, $\underline{y}$, are

obtained; however, these terms are probably already small for any reasonable calibration run, and to construct an additional credibility approximation for the posterior-to-calibration variance would require additional moments and complex formulae. Similarly, it should be possible in principle to revise our estimate of H(X) after the measurement y is made, but this leads to the same additional complexity. If one wishes, posterior to the calibration, one can replace b_2 in (5.1) by $g_2(\underline{y})$.

We mention again some of the limiting cases of (5.2)(5.3), assuming that the revised instrument covariances are small. First, if the observation error variance σ_M^2 is very large, or the variance in input is small, then the credibility in (5.3) will be very small, and the best estimate of the input is the prior mean. Conversely, a diffuse input, $\mathscr{V}\{\tilde{x}\} \to \infty$, will lead to $z_1(\underline{y},X) \approx (g_2(\underline{y}))^{-1}$, and a forecast:

$$f(y|\underline{y},X) \approx [y - g_1(\underline{y})]/g_2(\underline{y}) \quad . \tag{5.4}$$

6. Optimization

For the optimization, we assume that there is a total of T hours to be split among n calibration measurements, say a total of T_C hours, and the remainder, $T_M = T - T_C$ hours, to be spent upon r inverse inference measurements. We assume that one hour spent on a single measurement or calibration gives an observation error variance of σ^2; therefore the individual observation variances used previously are then:

$$\sigma_C^2 = \frac{n\sigma^2}{T_C} \quad ; \qquad \sigma_M^2 = \frac{r\sigma^2}{T_M} \quad . \tag{6.1}$$

To minimize the prior-to-calibration estimation of the forecast variance of a typical measurement, we must minimize the denominator of the second term of H(X) in (5.1):

$$D(T_C,T_M) = \frac{r\sigma^2}{T_M} + \mathscr{V}\{\tilde{x}\}(b_2^2 + \phi_{22}) + \phi_{11} + 2\phi_{22}\mathscr{E}\{\tilde{x}\} + \phi_{22}[\mathscr{E}\{\tilde{x}\}]^2 \quad . \tag{6.2}$$

where ϕ is given by (4.12), with σ_C^2 replaced by $n\sigma^2/T_C$ in (4.8), subject to $T_C + T_M = T$. In general, this optimization must be carried out numerically. However, if $n\sigma^2/T_C$ is much smaller than the diagonal terms of ΔM, then the calibration will have practically full credibility, and

$$\phi = (I_2 - \underline{z})\Delta \approx \left[I_2 - (I_2 - \frac{n\sigma^2}{T_C}(X'X)^{-1})\right] = \frac{\sigma^2}{T_C} M^{-1} \quad . \tag{6.3}$$

This shows the expected result, namely, that a good calibration run gives vanishing ϕ as T_C increases. The effect of the number of runs, n, is essentially cancelled out, as long as M is stable over different designs.

With this approximation, (6.2) can be written:

$$D(T_C,T_M) = \frac{r\sigma^2}{T_M} + \frac{\mu\sigma^2}{T_C} + \mathscr{V}\{\tilde{x}\}\ b_2^2 \quad , \tag{6.4}$$

where

$$\mu = \frac{m_2 - 2m_1\mathscr{E}\{\tilde{x}\} + \mathscr{E}\{\tilde{x}^2\}}{m_2 - m_1^2} \quad . \tag{6.5}$$

In this form, the optimization is obvious--the total time T should be split:

$$T_C^* / T_M^* = \sqrt{\mu/r} \quad , \tag{6.6}$$

giving a minimal value for D of:

$$D^* = \frac{\sigma^2}{T} (1 + \sqrt{\mu/r})^2 + \mathcal{E}\{\tilde{x}\} \, b_2^2 \quad . \tag{6.7}$$

An increase in the number of production runs, r, decreases the time used for calibration in an interesting way (6.6).

It is also interesting to note, in this approximation, that the ratio of effort depends, in addition to r, only on the first and second moments of the calibration design inputs, and on the measurement input. If the design X is considered to be variable, we see that we can further minimize (6.4) by decreasing μ, i.e. we choose inputs $\underline{x}$ so that:

$$m_1 \approx \mathcal{E}\{\tilde{x}\} \quad ; \quad (m_2 - m_1^2) \text{ is as large as possible;} \tag{6.8}$$

which is very intuitive from a physical point of view.
This design choice would make μ close to unity, and then $T_C^*/T_M^* = r^{-\frac{1}{2}}$. Of course, there may be many other physical reasons why the calibration input must be chosen in a different manner.
Even if the approximation (6.3) does not hold, (6.6) is suggested as an initial trial solution.

7. Numerical Example: Calorimetric Measurement of Nuclear Material

In order to illustrate the models developed in previous sections we use three kinds of information:

(1) a-priori information on the relationship between dependent and independent variable;

(2) results of calibration;

(3) results of measurement of the dependent variable.

The following realistic example will illustrate circumstances under which certain information is more important, and the improvement is achieved by using credibility procedures.

Let us consider the quantitative measurement of plutonium with the help of a *calorimeter*. The problem is to measure a voltage induced by the heat produced by the plutonium. For this purpose, one has to know the isotopic composition of the plutonium to be measured as well as the specific heat production of the different isotopes. Typical data are given in Table 1.

Let the amount of plutonium of one batch to be measured, and let w be the specific heat production of the plutonium under consideration. Then the heat x produced by the amount P of plutonium is given by

$$x = w \cdot P \quad . \tag{7.1}$$

The voltage E_M induced in the measurement chamber of the calorimeter is proportional to this heat:

$$E_M = a \cdot x = a \cdot (wP) \quad . \tag{7.2}$$

In a second, identical chamber, a reference heat x_0 is generated which induces a voltage E_0. Because of the assumed symmetry of the chambers, we have

$$E_0 = a \cdot x_0 \quad . \tag{7.3}$$

The value of x_0 is kept constant throughout the operation of the instrument. The quantity actually measured is the differential voltage y,

$$y = E_0 - E_M = a \cdot x_0 - a \cdot (wP) \quad ; \tag{7.4}$$

or, in other words,

$$y = \beta_1 + \beta_2 \cdot (wP) \quad , \tag{7.5a}$$

where

$$\beta_1 = a \cdot x_0 \quad , \qquad \beta_2 = -a \quad , \qquad a > 0 \quad . \tag{7.5b}$$

The value of x_0 may be assumed to be known precisely. In addition, we assume there exists experience from past measurements, expressed as expectation and variance of $\tilde{a}$, now considered as a random variable. This means we know

$$b_1 = \mathscr{E}\{a\}x_0 \quad ; \qquad b_2 = -\mathscr{E}\{a\} \quad ; \tag{7.6a}$$

$$\Delta = \begin{pmatrix} \mathscr{V}\{\tilde{\beta}_1\} & \mathscr{C}\{\tilde{\beta}_1 : \tilde{\beta}_2\} \\ \mathscr{C}\{\tilde{\beta}_1 : \tilde{\beta}_2\} & \mathscr{V}\{\tilde{\beta}_2\} \end{pmatrix} = \mathscr{V}\{\tilde{a}\} \begin{pmatrix} x_0^2 & -x_0 \\ -x_0 & 1 \end{pmatrix} \quad . \tag{7.6b}$$

The calibration is performed by putting an electric heater into

the measurement chamber and generating different values x_{i2} of heat which generates corresponding differential voltages y_i:

$$\Delta E_i = \beta_1 + \beta_2 \cdot x_{i2} \quad , \qquad i = 1,\ldots,n \quad . \tag{7.7}$$

Typical data for such a measurement problem are given in Table 2. According to this table, we have

$$b_1 = 600 \ [\text{mV}] \quad , \tag{7.8a}$$

$$b_2 = -240 \ [\text{mV/Watt}] \quad ; \tag{7.8b}$$

and furthermore,

$$\mathscr{V}\{\tilde{a}\} = (.02)^2 \cdot [\mathscr{E}(\tilde{a})]^2 = 23.04 \ \cdot \ [\text{mV}^2/\text{Watt}^2] . \tag{7.8c}$$

In addition, we have

$$\mathscr{E}\{\tilde{x}\} = 2.668, \quad \mathscr{V}\{\tilde{x}\} = .07118, \quad \mathscr{E}\{\tilde{x}^2\} = 7.189 \quad . \tag{7.9}$$

Therefore, we get for Δ_{ij}, as defined by (3.10) and given by (7.6),

$$\Delta = 23.04 \begin{pmatrix} 6.25 & -2.5 \\ -2.5 & 1 \end{pmatrix} = \begin{pmatrix} 144 & -57.6 \\ -57.6 & 23.04 \end{pmatrix} . \tag{7.10}$$

Let us consider first the case that we do not perform any calibration, but use only the prior information given by equations (7.8) and (7.9). According to (3.12) the estimate of the heat production is given by

$$f(y) = \mathcal{E}\{\tilde{x}\} + z_1(y - \mathcal{E}\{\tilde{y}\})$$

$$= 2.48 \cdot 10^{-3} + \frac{y - 600}{b_2 + 0.2234} \quad , \qquad (7.11)$$

which is to a good approximation

$$f(y) \sim \frac{1}{b_2} (y - 600) \quad .$$

We can easily determine the preposterior improvement in precision if we use (7.11) instead of simply using $\mathcal{E}\{\tilde{x}\}$; if we take $\mathcal{E}\{\tilde{x}\}$, then the variance of this estimate is

$$H_0 = \mathcal{V}\{\tilde{x}\} = .07118 \quad .$$

Now, according to (3.18) we get for the variance of the forecast error of a single measurement

$$H = \mathcal{V}\{\tilde{x}\} \cdot (1 - z_1 \cdot b_2)$$

$$= \mathcal{V}\{\tilde{x}\} \cdot 9.31 \cdot 10^{-4}$$

$$\approx 10^{-3} \cdot \mathcal{V}\{\tilde{x}\} \quad ,$$

and according to (3.19), for the variance of the forecast error of the sum of r measurements

$$H^{(r)} = r \cdot \mathcal{V}\{\tilde{x}\}(1 - z_1 \cdot b_2) + (r^2 - r) \cdot z_1^2 \cdot (\Delta_{11} + 2\Delta_{12} \cdot \mathcal{E}\{\tilde{x}\} + \Delta_{22} [\mathcal{E}\{\tilde{x}\}]^2)$$

$$= 4.3 \cdot 10^{-3} + 4 \cdot 10^{-2}$$

$$\approx 4.4 \cdot 10^{-2} \quad ,$$

which shows that this variance is mainly determined by the uncertainty of the instrument parameters, which is common to

all measurements.

Let us now use the calibration given in Table 2. With

$$X = \begin{pmatrix} 1 & .8 \\ 1 & 1.1 \\ \vdots & \vdots \\ 1 & 2.9 \end{pmatrix} \tag{7.12}$$

we have

$$X'X = 8 \begin{pmatrix} 1 & 1.85 \\ 1.85 & 3.845 \end{pmatrix} = 8 \cdot M \quad . \tag{7.13}$$

We can use the approximate formula (6.6) for the optimal distribution of calibration and measurement effort, if $n \cdot \sigma^2/T_c$ is much smaller than the diagonal terms of $\Delta \cdot M$. We check this assumption by first using equation (6.6) and then seeing whether or not the result fulfills the assumption.

According to equation (6.6) and Table 2 the optimal distribution of the time T available is given by

$$\frac{T_c^*}{T_M^*} = .214 \quad , \qquad T_c^* + T_M^* = 720 \quad ,$$

or, in other words,

$$T_c^* = 127 \quad , \qquad T_M^* = 593 \quad . \tag{7.14}$$

Therefore, we have

$$\sigma_c^2 = \frac{n \cdot \sigma^2}{T_c^*} = 1.154 \ll \begin{pmatrix} |(\Delta M)_{11}| \\ |(\Delta M)_{22}| \end{pmatrix} = \begin{pmatrix} 300 \\ 142 \end{pmatrix} \quad , \tag{7.15}$$

which means that our assumptions are fulfilled.

Finally, we want to determine the improvement in precision by using the calibration. According to equation (4.12) we have

$$\phi(X) = (I_2 - \underline{z}) \cdot \Delta \quad ,$$

where $\underline{z}$ is given by (4.8). With (7.10), (7.13), and (7.15) we obtain

$$\underline{z} = \begin{pmatrix} 5.96 & 12.54 \\ -2.36 & -4.94 \end{pmatrix} \quad ,$$

which gives for (4.12)

$$\phi = \begin{pmatrix} 8.06 & -3.22 \\ -2.34 & 0.96 \end{pmatrix} \quad . \tag{7.16}$$

Even though the forecast error variance after calibration and measurement according to (5.1) can be determined only if the calibration data (x_i, y_i), $i = 1,\ldots,n.$ are available, a comparison of (7.16) and (7.10) shows that the use of the calibration represents a considerable improvement in precision.

Table 1: Typical Plutonium Mixture

(Source: Schneider et al. [12])

	Pu238	Pu239	Pu240	Pu241	Pu242	Am241
Mean concentration [%]	0.041	90.51	8.265	1.113	0.064	0.05
Specific heat flux [mW/g]	569.0	1.923	7.03	4.62	0.12	108.4
Contribution to w [mW/g]	0.2333	1.7405	0.581	0.052	$7.69 \cdot 10^{-5}$	0.0612

Mean specific heat flux w: 2.668 [mW/g Pu]

Table 2: Typical Measurement Problem
(Source: Schneider et al. [12])

No. of batches r	60
Mean Pu content P [hg] of one batch	1
Mean heat production $x = w \cdot P$ [W] of one batch	2.668
Batch-to-batch variation	10%
Variance of a single measurement $\sigma^2(t)$ $[(mV)^2]$ as a function of time t[h] for $t > 6$	$\frac{18.324}{t}$
Total time T[h] available	720
No. of calibrations n	8
Range R of calibrations [Watt]	$0.8 \leq R \leq 3.0$
Values x_{i2} of calibration procedure	0.8, 1.1,...,2.9
A priori information $\mathscr{E}\beta_1$ [mV] on intercept β_1	600
A priori information $\mathscr{E}\beta_2$ [mV/Watt] on the slope of the calibration line	-240
A priori information on the variance of β (parametrically)	2%, 5%

References

[1] Brownlee, K.A., Statistical Theory and Methodology in Science and Engineering, Wiley, New York (1965).

[2] Cox, D.R., and Hinkley, D.V., Theoretical Statistics, Chapman and Hall, London (1974).

[3] Dunsmore, I.R., "A Bayesian Approach to Classification," Jour. Roy. Statist. Soc. (B), 28, pp. 568-577, (1966).

[4] Dunsmore, I.R., "A Bayesian Approach to Calibration," Jour. Roy. Statist. Soc. (B), 30, pp. 396-405, (1968).

[5] Hoadley, B. "A Bayesian Look at Inverse Regression," Jour. Amer. Statist. Assoc., 65, pp. 356-369, (1970).

[6] Jewell, W.S., "Exact Multidimensional Credibility," Mitteilungen schweizerischer Versicherungsmathematiker, 74, 2, pp. 194-214, (1974).

[7] Jewell, W.S., "Bayesian Regression and Credibility Theory," Internal Paper, IIASA, Laxenburg, Austria (March, 1975).

[8] Lindley, D.V., Bayesian Statistics, A Review, Regional Conference Series in Applied Mathematics, No. 2, SIAM, Philadelphia, (1972).

[9] Miller, R.G., Simultaneous Statistical Inference, McGraw Hill, New York, (1966).

[10] Muth, R.F., "The Demand for Non-Farm Housing," in The Demand for Durable Goods, A.C. Harberger (Ed.), The University of Chicago Press (1960).

[11] Press, S.J., and Scott, A., "Missing Variables in Bayesian Regression," in Studies in Bayesian Econometrics and Statistics, S.E. Fienberg and A. Zellner (Eds.), North-Holland, Amsterdam, pp. 259-272, (1974).

[12] Schneider, V.W., Hille, F., Kiy, M., and Gmelin, W. "Instruments Available for Safeguarding Fuel Fabrication Plants," Proceedings of a Symposium on Safeguards Techniques, Vol. I, pp. 181-200, International Atomic Energy Agency, Vienna, (1970).

[13] Zellner, A., "An Introduction to Bayesian Inference in Econometrics," Wiley, New York (1971).